Proceedings of the 2004 IEEE
Conference on Cybernetics and Intelligent Systems
Singapore, 1-3 December, 2004

Learning and Recognizing Behavioral Patterns Using Position and Posture of Human

Shigeki Aoki†* Masaki Onishi‡ Atsuhiro Kojima†† Kunio Fukunaga‡‡**
†Department of Electronic Control Engineering, Kumamoto National College of Technology, Kumamoto, Japan
‡Bio-Mimetic Control Research Center, RIKEN, Aichi, Japan
††Library and Science Information Center, Osaka Prefecture University, Osaka, Japan
‡‡Graduate School of Engineering, Osaka Prefecture University, Osaka, Japan
E-mail: *shigeki_aoki@m.ieice.org, **fukunaga@cs.osakafu-u.ac.jp

***Abstract*— In general, it is possible to find certain behavioral patterns in human daily activity. Such patterns are called as daily behavioral patterns. The purpose of this research is to learn and recognize behavioral patterns. In the previous methods, it is difficult to recognize in detail how a person acts in a room because the methods recognize only a sequence of existing position of human by using the information of infrared sensors or of switching on/off of electrical appliances. On the other hand, many have proposed the methods recognizing human motions from sequential images, in most of which motion models must be prepared in advance. In this paper, we propose a method for learning and recognizing motions of human without any motion models. In addition, we also propose perceptive methods of recognizing behavioral patterns by taking not only the sequence of position but also the sequence of motion into consideration. Experiments show that our approach is able to learn and recognize human behavior and confirm effectiveness of our method.**

I. Introduction

It is pointed out that there exist certain patterns in human behavior in general. We call such patterns behavioral patterns. This paper examines the possibility of learning and recognizing the behavioral patterns.

Many methods [1], [2], which observe behavioral patterns of a human mainly, have proposed as the use for monitoring system of the solitary senior citizen. In these methods, the daily behavioral patterns of the human are acquired by making use of the position obtained from the infrared sensors, which are installed in the living room, the bathroom and the rest room etc. and information of on/off of the electrical appliances. These methods being to handle the information obtained by specific switches at regular intervals could detect the state of the human well. It is hard, however, to examine what kinds of human motions appear in a room.

On the other hand, many have proposed the methods of recognizing human motions from sequential images. Bobick, Yamato and Wilson can be listed as representative methods [3], [4], [5]. In these methods, motion is recognized by comparing the input images with the movement of the human in the models, which are given beforehand. In the case when these methods are used under general environment, it is necessary to prepare many models, which are adjusted to the change of environment and personal equation. In addition, although independent motion is recognized, these could not consider order of motions.

In this paper, we propose a method for learning and recognizing human motions without any motion models, and also propose a method for recognizing the behavioral patterns by taking the motion order into consideration. In our method, first, information of position and attitude of the human is extracted from input images obtained from an omni-directional camera. In the next, sequences of the features are converted to the strings of symbols. Those strings are classified into certain number of motion types automatically. After that, our system identifies the motion when motion of the person is observed. In addition, our system converts behavioral patterns, which are represented by sequences of motions to the strings, learns and recognizes in the same way. Since our method is possible to examine the state of the person in detail without preparing the many models, it is easy to apply to the monitoring system for the solitary senior citizen.

II. Learning human motions

For the purpose of watching and seeing human activities, it is desirable for the sensor to watch the information of wide range at once. For this reason, we use omni-directional camera, which is capable to catch the scene of around 360 degrees at once. In order to recognize the detailed motions of the human, it is important to extract the position of the human, furthermore position relationship of the hand and the face. The body and the skin regions of the human are extracted from input images captured by the omni-directional camera. Using the geometry of these regions, the feature quantities which will be defined below are obtained.

A. Expressing human motions by strings

Firstly we extracted the moving region in the input image obtained from omni-directional camera, then execute the processing of expansion-contraction for noise removal, and the processing of isolated point removal twice at a time. Considering the case when more than one person appear in the image, we extract the human regions by making the moving regions cluster and approximate them as ellipse regions. We define seven feature quantities as shown in Table I for characterizing

TABLE I

FEATURE QUANTITIES FOR HUMAN MOTIONS.

feature quantity	corresponding feature	number of dimentions
COG of moving regions	person's position	2
pixels of moving regions	degree of motion	1
distribution of regions	person's pose	4
COG of skin regions	face and hands postions	2
pixels of skin regions	person's direction	1
distribution of skin regions	relative position of skins	4
number of persons	number of persons	1

human motions. For example, the center of gravity (COG), the number of pixels, and the distribution of the moving regions indicate position and attitude of the human.

The feature extracted with above method is expressed by a vector with 15 dimensions. The dimensions of these features are reduced by making use of eigenspace method. In the next, after clustering on the eigenspace, each cluster is labeled. The system projects the feature parameters on the eigenspace when motion of the human is observed, and examines a Mahalanobis distance from each of clustering. After obtaining the label of clustering which is shortest from the point of observed motion, the system represents the motion of human by the string of labels. Here, we assume that the duration of one motion is from the start time of extracting feature parameters to the time of disappearing of them.

B. Learning and recognizing motions

Using the method for representing the motion by the string of labels described above, the system produces the similar string in the case when the similar motion appeared. For example, similar strings are observed in the case when a person moves from a bed to a table many times. But the different strings appear when the person moves from the table to a bathroom. In the case when let all the strings apply a HMM (Hidden Markov Model), the likelihood gets high even if the different string from the learning strings is given. To prevent this kind of problem, it is necessary to prepare the same number of HMMs as the number of different motions. It is difficult to know the number of different motions in advance. In our approach, we classify the motions using the likelihood of given number of HMMs.

Learning algorithm of motion is shown as follows.

[An algorithm for learning motions.]

(Step1) Compose a HMM using all of the observed strings as the learning strings.

(Step2) Obtain all of the likelihood of learning strings for a composed HMM. Plotting these data on this axis we obtain a histogram.

(Step3) Search for the minimum value around the standard value, which is calculated by applying discriminant analysis to the distribution of histogram, and let it be the threshold.

(Step4) The set of strings is classified into 2 types by comparing the threshold value with each likelihood of strings. And, the flag of "0" is attached to the set of strings under the threshold value, the flag of "1" is attached to the set of strings above the threshold value.

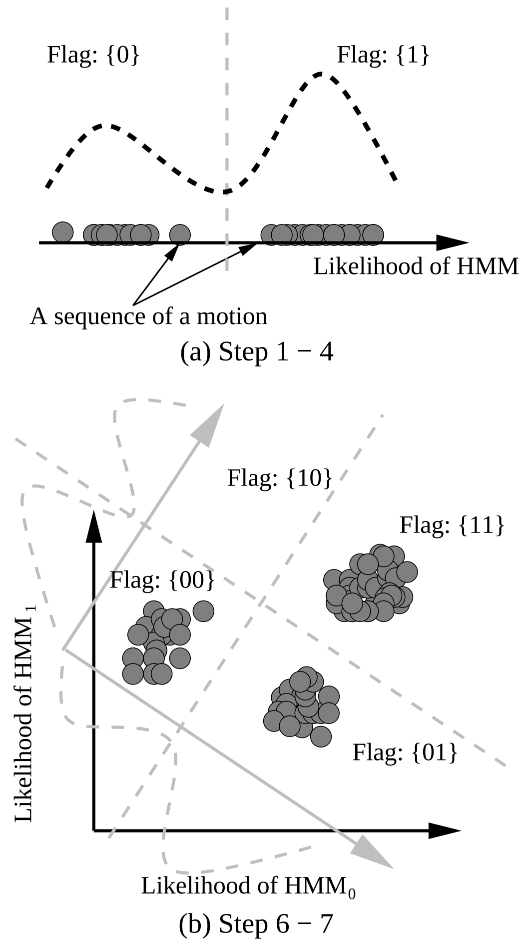

Fig. 1. An algorithm of learning motions.

Figure 1 (a) shows the flow of the processing from **(Step1)** to **(Step4)**. The circle in the figure shows the string, which corresponds to one motion. It is plotted with the likelihood of HMM as the axis. The dotted line of black shows the histogram, and the gray dotted line shows the threshold value, which is calculated with **(Step3)**. In addition, Flag:$\{0\}$, Flag:$\{1\}$ shown in the figure express the flag, which is attached to the set of strings. The set of motions appeared during learning period can be divided into 2 types by the processing described above.

(Step5) Compose the HMM using strings included in the set, which is the one of sets.

(Step6) Calculate the likelihood of all strings. All the strings are projected to the space where the each axis is given by the likelihood of each HMM, and obtain the basis vector applying PCA (Principal Composed Analysis) to the space obtained above.

(Step7) Compose the sub-space using the basis axis of which rate of cumulative contribution exceeds threshold value. Calculate the threshold values of each basis axis same as **(Step3)**, the flag is attached to each string.

(Step8) Iterate the **(Step5)** through **(Step7)** until the set of strings does not differ from previous loop processing.

Figure 1 (b) shows the flow of the processing at the 1st loop. The circle shows the string of motion, and is plotted on the 2-dimension plane of which axes show the likelihood of two HMMs. The axis shown by gray line expresses the basis vector obtained by PCA, and the flag of each string is decided by comparing with the threshold of each axis. Then the figure shows that the motions during learning period are classified into 4 types (Flag:$\{00\}$, Flag:$\{01\}$, Flag:$\{10\}$, Flag$\{11\}$). It is difficult to know the number of motion types in advance. In other word, it is hard to know what number classification of the motions is suitable in advance. Therefore iterate **(Step6)** and **(Step7)** until condition of **(Step8)** is satisfied. Increase the number of axes of classification space by increasing the number of HMMs at each step. With this processing, it is possible to classify the motions robustly.

Retain the HMMs, the basis vectors, threshold values, which are calculated at above processing and the flag (for example in Figure1 (b), all sets other than $\{10\}$) of which set contains the string more than threshold.

To recognize the motion, the system calculates the likelihood of the input string for all HMMs, and then obtains the flag using the flag space with the basis axis. Comparing this flag with the flags of motions, the system recognizes the motion corresponding to the string. The system decides to be non-daily motion in the case when the input flag does not correspond to any retained flags.

III. Learning and Recognizing Behavioral Patterns

Section **II** describes the recognition method for the independent motion such as "The human walked from the bed to the refrigerator." and "The human walked from the refrigerator to the table." etc. It can be considered that there is a relationship between motions, which are done within short time. For the example mentioned above, it is natural to think "The human moves to take the food and beverage in the refrigerator, and has the meal at the table." In addition, it can be considered that the motions within short time are regular and habitual motions. This section describes the method for learning and recognizing behavioral patterns considering the order of motions.

First, in order to express behavioral patterns by string, the system labels each motion. In the next, sequence of observed motions is converted to the string of behavioral pattern. The system learns all of strings given by observed behavioral patterns on the basis of the algorithm described in section **II-B**. Each flag obtained by learning corresponds to a behavioral pattern of human.

Also, the method of section **II-B** is applied when recognizing which kind of input behavioral pattern is. The system obtains the flag of the observed behavioral pattern. The input behavioral pattern can be decided to be the pattern with the some flag. When the same flag does not exist, input pattern is considered non-daily behavioral pattern.

TABLE II

Results of learning.

class	resulted flags	corresponding motions				
		A_1	A_2	A_3	$\cdots$	A_8
a_1	$\{00100011111011\}$	24				
a_2	$\{00110011111011\}$	3				
a_3	$\{00100011111111\}$	1				
a_4	$\{00100011111101\}$	1				
a_5	$\{10100011111011\}$	1				
a_6	$\{10100101101011\}$		24			
a_7	$\{10100001101011\}$		1			
a_8	$\{10100001101111\}$		1			
a_9	$\{00100111111001\}$		1			
a_{10}	$\{00000011100001\}$			29		
a_{11}	$\{00000011101001\}$			1		
$\vdots$		$\vdots$				
a_{26}	$\{10110101101011\}$		3			12
a_{27}	$\{10110101100011\}$					15
a_{28}	$\{00110101101011\}$					1
a_{29}	$\{10100101101111\}$					1
a_{30}	$\{10110101100010\}$					1

TABLE III

Results of recognition of motions.

class	results of recognition				
	A_1	A_2	A_3	$\cdots$	A_8
a_1	15				
a_2	1				
a_3	2				
a_4					
a_5					
a_6		15			**1**
a_7					
a_8					
a_9					
a_{10}			20		
a_{11}					
$\vdots$	$\vdots$				
a_{26}		1			9
a_{27}					9
a_{28}					1
a_{29}					
a_{30}					
non-daily	**2**	**4**		**3**	

IV. Experiments and Discussions

To examine our approach, we made an experiment on learning and recognizing the human motions, and the behavioral patterns. Images are taken from omni-directional camera, which is placed the center of the laboratory with the resolution of 320 $\times$ 240 pixels. In the first, we experimented in regard to the learning and recognizing motions. 8 types of motions ($A_1 \sim A_8$), which are shown in Figure 2, are used for learning and recognizing experiment. The white curve line in the figure shows the trajectory of position of human. However, scene A_2 and scene A_8 shows the same trajectory except different attitude.

30 times of motions for each scene were done, and the system used them as learning motions. As results, the motions are classified into 30 types, which are shown in Table II. Another 20 motions for each scene were used for the

Fig. 2. Test motions used in experiments.

experiment of recognition. Table III shows the results. As shown in the table, it was possible to identify 150 strings (motions) out of 160 strings (motions). The rate of correct recognition is 93.5%. Since omni-directional camera is far from the persons, 10 strings, which are marked as bold in Table III, had failed to extract the feature quantity. In addition, we added the experiments to examine non-daily motion. We prepared to motions for each of another 5 types of scene. As results of experiments, all strings (motions) could identify the non-daily motion.

In the next, we experimented in regard to learning and recognizing behavioral patterns. 7 consecutive motions of 8 types, which are chosen as motion experiments, were used in this experiment. 30 strings were composed by the 10 patterns for each of 3 behavioral patterns ($B_1 \sim B_3$). In addition, we prepare 6 patterns for each of behavioral patterns, which are used for recognition. The results are shown in Table IV. The system learned 30 strings (behavioral patterns) for learning then, identified the 18 strings (behavioral patterns) for recognition. As the results, the system identified 17 patterns correctly. The rate of correct recognition was 94.4%. Furthermore, we did the experiment, which identifies non-daily behavioral pattern. We prepared to 6 behavioral patterns for each another 2 types of scene. As the result of experiment for those strings (behavioral patterns), all strings can be identified non-daily pattern.

V. Conclusion

In this paper, we proposed a method for learning and recognizing human motions without any motion models. In addition, we also propose a perceptive method for recognizing the behavioral patterns by taking the motion order into consideration. The experiment confirms effectiveness of our approach. As our method is possible to know the behavioral state of the person in detail without any models, it is easy to apply our approach to the monitoring system for the solitary senior citizen.

TABLE IV

Results of learning and recognizing behavioral patterns.

class	resulted flags	learning patterns (recognized patterns)		
		B_1	B_2	B_3
b_1	{010110}	6(2)		
b_2	{110100}	4(4)		
b_3	{111001}		2	
b_4	{101001}		4(3)	
b_5	{110001}		2	
b_6	{100001}		1(1)	
b_7	{101011}		1(1)	
b_8	{110101}			10(6)
non-daily			**(1)**	

Acknowledgments: We are grateful to Mr. Yoshio Iwai for helping experiments.

References

[1] S. Ohta, H. Nakamoto, Y. Shinagawa, T. Tanikawa, “A health monitoring system for elderly people living alone,” Journal of Telemedicine and Telecare, 8, pp.151–156, 2002.

[2] T. Matsumoto, Y. Shimada, K. Shibasato, H. Ohtsuka, and S. Kawaji, “Creating behavior model of a senior solitary life and detecting an urgency,” Proc. International ICSC Congress on Intelligent Systems & Applications (ISA2000), vol.2, pp.234–238, 2000.

[3] A. F. Bobick, and J. W. Davis, “The recognition of human movement using temporal templates,” IEEE Trans. Pattern Anal. & Mach. Intell., vol.23, no.3, pp.257–267, 2001.

[4] J. Yamamoto, J, Ohya, and K. Ishii, “Recognition human action in time-sequential images using hidden markov model,” Proc. IEEE Conf. Computer Vision and Pattern Recognition, pp.379–385, 1992.

[5] A.D. Wilson, and A.F. Bobick, “Parametric hidden Markov models for gesture recognition,” IEEE Trans. Pattern Anal. & Mach. Intell., vol.21, no.9, pp.884–900, Sept. 1999.