Care Giving System Based on Consciousness Recognition

Noriko Yokoyama, Tomoyuki Yamaguchi, and Shuji Hashimoto

Department of applied physics, Waseda University,
3-4-1 Okubo, Shinjuku-ku, Tokyo, Japan
yokoyama@shalab.phys.waseda.ac.jp

Abstract. In these days, robotics systems that provide supportive communication to human have been actively developed. However in such systems the internal consciousness state of human is not taken into consideration and hence the provision of support might not be appropriate. In this article we proposed a study support communication system that encourages and praises the human user based on the recognition of consciousness state through the user posture.

Keywords: Human interface, Bayesian network.

1 Introduction

The robots that support human life such as a humanoid robot and a pet robot have been actively developed. These robots have to observe human behavior to provide appropriate supports. In such a system, if the human internal state is recognized through the person's behavior, the robot will be able to offer more suitable support to the people. Moreover, while the development of the robot that supports people's daily life is advanced, it is pointed out that we do not feel like the robot blending into society [1]. To develop a more familiar robot, we should construct a relation between people and a robot [2][3] and communication between people and robots is important. The essence of communication is the exchange and sharing of state of mind [4]. It is necessary that we provide a robot with an internal state of human and a function that makes the robot seems like having mind and it is important how a robot extracts internal state of human from this point of view.

As one of the systems that observe people to support them, there is a cooperative system of the environmental type robot and the individual robot, where the location information of people in the house is detected by pressure sensors spread over the floor and infrared cameras. The location information is used for an appliance operation [5] or a physically support like taking something [6]. These systems can support people, but the internal state of the people is not considered. While, as the research that estimates the internal state of people, there is a sleep detection system while driving [7]. The consciousness level is estimated by detecting the frequency and opening time of blinks using images captured with the infrared camera for warning.

In this article, we developed a system that provides supportive communication based on consciousness recognition of the user. We show the details of the proposed system and the experimental results.

M.J. Smith, G. Salvendy (Eds.): Human Interface, Part I, HCII 2011, LNCS 6771, pp. 659–668, 2011.

2 Proposed System

The proposed system is composed of three parts as shown in Figure 1. First, the images of a user are taken with a USB camera and the posture of the user is estimated. Secondly, the user's state of consciousness is estimated from the time-sequence data of the posture using neural network. Finally, the system selects an utterance and utters to the user based on the time-sequence of the state of consciousness using Bayesian network.

2.1 Posture Estimation

We extract the silhouette of the user from an image using background difference method and estimate the posture using template matching. We prepared four types of postures "normal (upright) ", "leaning", "downward" and "face down ".

(1) Acquisition of Silhouette. User's silhouette is extracted by background difference method of gray scale image. The size of input image is 640×480 pixels. The library function of OpenCV is used for making grayscale image. The pixel value of (x, y) of pre-processed image is defined as f(x, y) and pixel value of (x, y) of post-processed image is defined as g(x, y) and we obtain silhouette image using Equation 1 where the threshold value is defined as T.

$$g(x, y) = \begin{cases} 255 & f(x, y) > T \\ 0 & f(x, y) \leq T \end{cases} \tag{1}$$

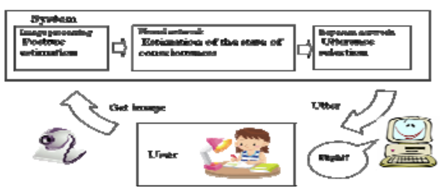

Fig. 1. Overview of system

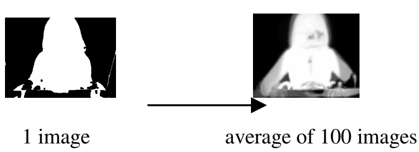

Fig. 2. Example of typical image

(2) Generation of Template Image. Four kinds of typical images are generated by averaging each pixel value on g(x, y) of 100 images of each posture group. These average images are used in next template matching. Figure 2 shows the example of typical image and the template image of "normal" obtained by averaging.

(3) Template Matching. The distances between silhouettes of input image P(x, y) of the user and each template image $Q_i(x, y)$ are calculated by Equation 2. The template image that has the smallest distance is detected. Its posture is estimated as the posture of the user.

$$d = \frac{1}{N_W} \sum_{(x,y)\in W} \left| P(x,y) - Q_i(x,y) \right| , \tag{2}$$

where i denotes the posture number.

To prevent the influence of the input situation on the desk, the range of silhouette W is defined as 640×420 pixels. N_W is the number of pixels in area W.

2.2 Estimation of State of Consciousness

Above-mentioned posture estimation is carried out at 2fps and the state of consciousness is estimated based on the temporal feature of the posture changes. As an input, we converted five minutes of posture data (length 600) into ten parameters composed of the total number of each posture:4 and the number of posture transition:6. As an output, we prepared four levels of consciousness state, ”concentrating”, ”sleepy”, ”lack the drive” and ”sleeping”. These levels were selected based on phase theory [8].

(1) Phase Theory. Phase theory was introduced by Kunie Hashimoto in 1978. This theory divides the consciousness state into five levels from phase 0 to phase Ⅳ (Table 1). In this research, four consciousness state (Table 2) was prepared referring to above-mentioned phase 0 to Ⅲ.

(2) Neural network learning. The learning of each state of consciousness is performed with a three-layered perceptron with back propagation algorithm (Figure 3). In this research, neural network was chosen because the pattern of postures are classified into either of four states of consciousness. The reason we used posture data with time width is that consciousness state can be read from continuous actions. Time length is five minutes in this research because too long length cannot correspond to the change of the state of consciousness.

Table 1. Stage of consciousness state

Phase	Mode of consciousness	Physiology
0	unconscious, trance	sleeping, brain attack
I	Subnormal	fatigue, doze ,drunken
II	normal, relaxed	Relaxed
III	normal ,clear	Active
IV	hypernormal, excited	Panic

Table 2. Consciousness states

Phase theory		Consciousness states
0 : sleeping	→	「sleeping」
I : doze	→	「sleepy」
II : relaxed	→	「lack the drive」
III : active	→	「concentrating」

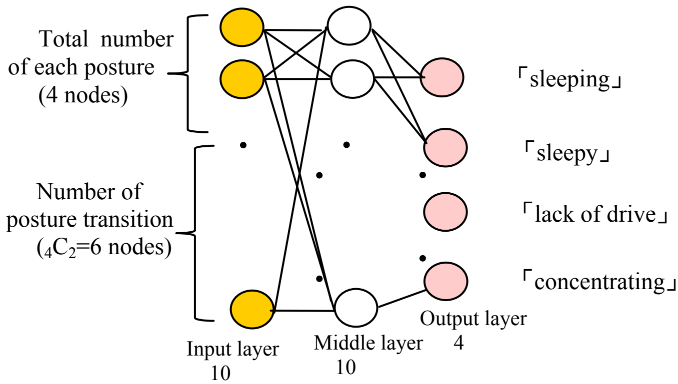

Fig. 3. Neural network

Table 3. Values of output

Consciousness states	Value of output
sleeping	(1,0,0,0)
sleepy	(0,1,0,0)
lack of drive	(0,0,1,0)
concentrating	(0,0,0,1)

As an input, we converted five minutes of posture data (length 600) into ten parameters composed of the total number of each posture and the number of posture transition. We use not only total number but also transition frequency between each posture, so time series information can be considered.

The teacher data is made by providing a state of five minutes of posture data and we let neural network learn the pattern. The output of each state is given as shown in Table 3. The learning of each state of consciousness is performed 10000 times. The average of error compared with learning data is

$$E = 3.734e^{-5} \fallingdotseq 0.025. \tag{3}$$

(3) Estimation of state of consciousness. We input converted ten parameters into learned neural network and the highest output is the estimated state at the time.

2.3 Utterance Selection

The system selects an utterance according to the time-sequence of consciousness using Bayesian network. Figure 4 shows the model of Bayesian network. There are 9 situation nodes and 10 utterance nodes. Table 4 shows situation nodes. These are calculated based on the time-sequence consciousness. For example, if the user follows the following path, the situation nodes at the time are calculated as shown in Table 5.

concentrating (30 minutes) → sleepy (15 minutes) → concentrating (20 minutes) → present

In learning phase, the system constructs the timing-structure model of human speech from the learning data. t_k is the value of a situation node s_j when a utterance c_i is observed. n is the total number of each utterance. The conditional probability $P_{s_j|c_i}(t)$ for an utterance node and a situation node are calculated as follows.

$$f_k(t) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(t-t_k)^2}{2\sigma^2} \right) \tag{4}$$

$$P_{s_j|c_i}(t) = \frac{\sum_{k=1}^{n} f_k(t)}{n} \tag{5}$$

In selecting phase, the probability of each utterance node is calculated using Equation 6 and the system utters a word when the probability exceeds the certain threshold and attains the local maximum point.

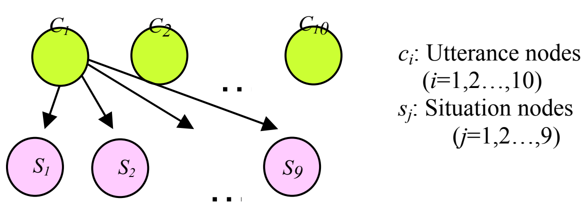

Fig. 4. Model of Bayesian network

Table 4. Situation nodes

The meaning of nodes	Number of nodes
Total time of a day of each state of consciousness	4
Duration of each state of consciousness	4
Elapsed time from the last utterance	1

Table 5. Example of situation nodes

Total time		Duration	
concentrating	: 50 minutes	concentrating	: 20 minutes
lack of drive	: 0 minutes	lack of drive	: 0 minutes
sleepy	: 15 minutes	sleepy	: 0 minutes
sleeping	: 0 minutes	sleeping	: 0 minutes

$$BEL(c_i) = P(S^i \mid c_i) = \prod_{j=!}^{m} P_{s_j \mid c_i}(s_j{}^i) \; . \tag{6}$$

Each threshold is set to 1/2 of the minimum BEL in study data.

3 Evaluation Experiments

We conducted evaluation experiments for each part and the whole system.

3.1 Experiment of Posture Estimation

First, we conducted the experiment of posture estimation. We supposed that we can classify postures into four postures and we confirmed whether the system can estimate the posture.

Posture images are taken for ten minutes with 2 fps (1200 frames) and judged the posture per one frame. We prepared four types of postures "normal (upright) ", "leaning", "downward" and "face down ". To get answers, posture images are preserved and the user judges it. Posture estimation from the system is compared with the correct answer and recognition accuracy is calculated. This experiment is conducted at a desk in the laboratory and the USB camera is set on the personal computer.

The threshold value of Equation 1 is set to 20 from preliminary experiments. If we set a smaller threshold, image is influenced by the flicker of the fluorescent lamp of the room. If we set a larger threshold, colors in background cannot be recognized.

Results of the accuracy of posture estimation were as shown in Table 6. Results showed high recognition rates of each posture but the rates of "downward" and "face down" are lower than that of the other postures. This is because downward or face down posture leaning a little is occasionally estimated as leaning posture.

3.2 Experiment of Consciousness Estimation

We conducted the experiment of consciousness estimation.

A user's activity data is taken for 7 days and a system estimates the user's consciousness per 5 minutes. Four levels of consciousness, "concentrating", "sleepy", "lack the drive" and "sleeping" are prepared. The estimation is compared with the correct answer that the user judges.

As a result, the accuracy of consciousness estimation was calculated as shown in Table 7. Results showed high recognition rates of each consciousness. However, when feeling a little tired even if concentrating, the state is sometimes estimated as lack of drive. Moreover, it was difficult to distinguish between “sleepy” and “lack of drive”. It was turned out that there were a few neutral states like this. To cope with these situations, it is one way that we make the system recognize that the judgment is inaccurate and prepare utterances corresponding to the situation.

Table 6. Accuracy of posture estimation

	Normal	Leaning	Downward	Face down	Total
Recognized frames (frames)	277	380	193	267	1117
Total number (frames)	287	382	229	302	1200
Accuracy (%)	96.5	99.5	84.3	88.4	93.1

Table 7. Accuracy of consciousness estimation

	Concentrating	Lack of drive	Sleepy	Sleeping	Total
Recognized number	146	9	10	10	175
Total number	148	9	11	10	178
Accuracy (%)	98.6	100	90.9	100	98.3

Table 8. Threshold of each utterance

	Utterance	Threshold($\times 10^{-13}$)
1	Are you ok?	0.098548
2	Wake up!	0.063544
3	You seem tired.	0.105729
4	Snap out of it.	0.038767
5	Do you want to sleep?	0.052691
6	Are you tired?	0.122829
7	Why don't you take a rest for a while.	0.292587
8	You started studying hard.	0.222411
9	You study hard!	0.002201
10	Don't work too hard.	1.279856

Table 9. Example of utterance corresponding to consciousness estimation (self-judge)

Consciousness state	Utterance from system
Sleeping	Are you ok?
Concentrating	You study hard!
Sleepy	Wake up!

3.3 Experiment of Utterance

We conducted the experiment of utterance part.

First, the teacher data are made by providing a teacher utterance for active data or unreal data and Bayesian network learns the utterance patterns. 67 data were used for its learning. Secondly, we calculated probabilities and decided thresholds of each utterance. Finally, we let users study at a desk while this system is running and observed utterances from the system. Software “Easy Speech” [9] is used for generating utterances.

We decided the threshold of each utterance as shown in Table 8. Each threshold is set to 1/2 of the minimum BEL in study data. Using these thresholds, we let the system running, and confirmed utterances as shown in Table 9. When concentrating, the transition of probability of “You study hard ! ” is as shown in Figure 5.

The threshold varied widely per utterance. This is because the probability of data far from other data becomes very small when the data increases and leans. For example, concerning utterance 9 “You study hard ! ” which has a lot of learning data, figure 6 shows the conditional probability of “Utterance 9” and “Duration of concentrating”. The probability after 100 minutes becomes extremely small where the number of teacher data is few. Therefore, it is difficult to use only threshold to judge an utterance timing and we put local minimum condition to the judgment method. Although the threshold of utterance 9 is 0.002201 and this is very small, it is expected that utterance is generated in both the first half part (35, 70, 100 minutes) when the probability is large and the latter part (140, 180, 220 minutes) when the probability is small.

Secondly, the utterances from the system were compared with consciousness state (self- judge) as shown in Table 9. The results show that the utterances are suitable to the situations like worry in case of sleeping and praise in case of concentrating. We also analyzed Bayesian probability. Figure 5 shows a time evolution of Bayesian probability. The probability of “You study hard!” increased gradually as concentrating time passed and the system uttered at its top. The reason of decrease after the last utterance is that passed time after utterance is one of situation nodes. The conditional probability between utterance 9 and passed time after the last utterance is as shown in Figure 7.

3.4 Effectiveness of Whole System

We confirmed the effectiveness of this system in daily life.

While this system is running, we let user lead a daily life in laboratory and observed utterances of the system. When the total of pixel values of user’s silhouette is less than 640*420*0.1=26880, it is judged that the user is not there and the system stops temporally.

Two people which are 1 female and 1 male participated in this experiment respectively for three days. The system was made by female’s teacher data. Table 10 shows the example of utterances. As a result, although there were some cases that no utterance was generated when long time passed, utterances were observed in any case within one hour. Moreover, we conducted a survey to the user and confirmed that the utterances were generated in appropriate circumstances.

The lack of the teacher data is thought as a reason why any utterance is not generated when long time passes. The variation of the transition increases as time passes, so the distance from teacher data becomes larger. To solve this problem, we should increase teacher data or make the variable of Bayesian network larger as time passes.

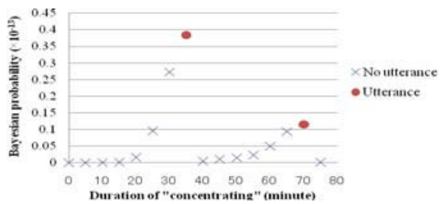

Fig. 5. Time-sequence data of Bayesian probability of "You study hard!" while concentrating

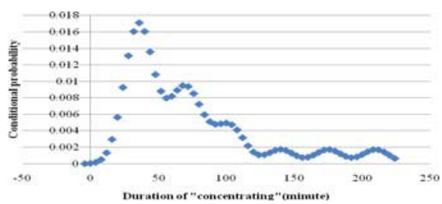

Fig. 6. Conditional probability of "Utterance 9" and "Duration of concentrating"

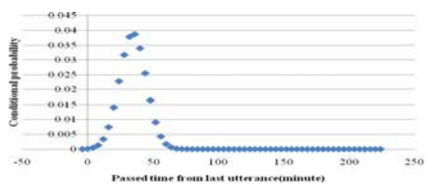

Fig. 7. Conditional probability of "Utterance9" and "Passed time from last utterance"

Table 10. Example of experimental result

Passed time（minute）	Consciousness state (system)	utterance
5	Sleepy	
10	Sleepy	
15	Sleepy	Wake up !
20	Lack of drive	
25	Lack of drive	
30	Concentrating	
35	Concentrating	You started studying hard!
40	Concentrating	
45	Concentrating	
50	Concentrating	
55	Concentrating	
60	Concentrating	
65	Sleepy	Wake up !

4 Conclusion

We developed a study support system as one of the spontaneous support system. This system got the awareness from user's appearance and talked to the user actively. We will try to focus on other motions and widen the scope of utterances in the future.

References

1. Okada, M.: Come Close to Me!: To be more Sociable Robots for Us. Journal of Human Interface Society 10(2), 87–92 (2008)
2. Oka, N., Kushi, K., Sonoyama, T., Shibuya, Y.: An Education in Interaction Design with Robot Materials. Journal of Human Interface Society 10(2), 105–108 (2008)
3. Ono, T., Ogawa, K.: Ubiquitous Mind and Interaction Design between Human and Robot. Journal of Human Interface Society 10(2), 115–118 (2008)
4. Kozima, H.: What Children Experience in Interactive Play with Robots. Journal of Human Interface Society 10(2), 99–104 (2008)
5. Ueda, H.: Human-Symbiotic Robots in the Home Ubiquitous Environment. Journal of Human Interface Society 10(2), 93–98 (2008)
6. Sato, T.: Networked Intelligent Machine to Support Daily Life —Trials of Robotic Room. Toshiba Review 56(9), 28–32 (2001)
7. Adachi, K.: Monitoring Car Drivers' Condition Using Image Processing. The Transactions of the Institute of Electrical Engineers of Japan. A Publication of Sensors and Micromachines Society 126(2), 31–37 (2006)
8. Hashimoto, K.: "Safety Human Engineering", Japan Industrial Safety & Health Association (1984) (in Japanese)
9. Easy speech, `http://www.vector.co.jp/soft/dl/winnt/art/se367527.html`