# RECOGNITION OF BASIC ACTIONS FOR THE DETECTION OF THE PERPLEX SITUATIONS USING BOTH PUPIL POSITIONS

Takayuki KAMITANI and Yoji MARUTANI
Assistant Professor, Professor
Department of Information Systems Engineering
Osaka Sangyo University
Nakagaito 3-1-1, Daito-shi, Osaka, 574-8530
E-mail: kamitani@ise.osaka-sandai.ac.jp, marutani@ise.osaka-sandai.ac.jp
JAPAN

Comission V, Working Group 4

**KEY WORDS:** DP Matching, Image Sequence, Motion Analysis, Perplex Situation, Perplexed Behavior

## ABSTRACT

This paper presents a method for recognizing the perplex situations based on the detection of basic actions by obtaining positions and an angle of pupils in the image sequence.
Firstly, the perplexed behaviors were picked up by observing the subjects in word processor work. As the result of the observation, it was made clear that the perplexed behaviors were shown in the motion of the head and the keyboard operation rather than the facial expressions. The typical head motions in the perplex situations are such as shaking, tilting, bending backward, bending forward to gaze on the CRT display, keeping still and changing the glance between the CRT display and the keyboard frequently. Secondly, the head motions in $x$, $y$, $z$ and $\theta$ directions were captured in real time by chasing both pupils through the use of image processing system. In the system, the processing speed is enhanced by reducing image data. The behaviors are converted to the series of vectors which have the moving velocity of head in $x$, $y$, $z$ and $\theta$ directions as the elements. Thirdly, the distances between unknown input motion patterns and the template patterns of the vector sequences were calculated by DP matching. As the result of DP matching, it was made clear that each head motion was recognized. The proposed method is available for not only the specified person but also unspecified persons. The method can be applied to the development of the software which responses automatically when the operator falls into the perplex situations.

## 1. INTRODUCTION

Now, we live in a computer assisted society and can't separate the computers from our life. However, even the computer which seems convenient never can be handled easily. Especially beginners feel difficulty in operating. With the coming of the age when everyone has to handle the computer, computer allergy has come to spread. For these reasons, it is desirable to make the computer carry out the kindness action to the operator who has fallen into the perplex situations. According to Ochiai et al. (Ochiai et al., 1994), few operators demand positively the computer a help for escaping the perplex situations. It is necessary to make the computer detect the operator's perplex situations. The previous approaches of mind state recognition have been done by the use of the facial expression analysis (Choi et al., 1991, Kitamura et al., 1993, and Matsuno et al., 1993). These approaches have been based on the theory that the facial muscles expand and contract corresponding to the state of mind. However, the facial expression analysis requires the processing of huge volumes of the CG data. Also, there is the problem that the analysis is extremely difficult when the face isn't toward the camera. For these reasons, we researched the simple and practical method of recognizing the perplex situations in real time by image processing (Kamitani and Marutani, 1994- 97e).

In this paper, we present a method for recognizing the perplex situations based on the detection of perplexed behaviors by obtaining positions and an angle of both pupils in the image sequence. In Sect. 2, we describe the observation of the perplexed behaviors. In Sect. 3, we propose the method of recognizing the basic actions of the perplexed behaviors using image processing and DP matching.

## 2. OBSERVATION OF PERPLEXED BEHAVIORS

### 2.1 Hardware configuration and observation method

To investigate the operator's perplexed behaviors, the subjects being at word processor work are observed. Figure 1 illustrates the setup for observing the perplexed behaviors.

The subject's face, the front and the side of his body, and a picture plane of the CRT display (PC-KM153 : NEC Co.) were observed by 4 television cameras (CCD-TR303 : Sony Co.), while he was doing the word processor work by the personal computer (PC-9821Ap : NEC Co.). The subject handles the word processor software (Ichitaro Ver.4 : JUSTSYSTEM Co.) which operates on MS-DOS (Microsoft Co.). Camera 1 is placed on the CRT display and captures his front-view face. Camera 2 captures upper half of his body from the front half right. Camera 3 captures the upper half of his body from the right side. Camera 4 captures the picture plane of the CRT display from the front half left. The images and the sounds of

each camera are recorded on the video tapes. The subjects are 31 persons who are unskilled to the personal computer and a word processor. It is easy to distinguish the perplex situations by observing facial expression, utterance, behavior and keyboard operation. We picked up the perplexed behaviors and counted the number of the subjects who took those actions by observing the replayed image sequence.

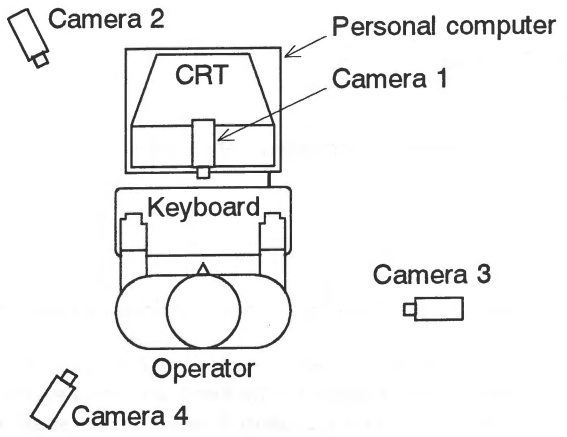



**Figure 1 Experimental system for observing perplexed behaviors.**

### 2.2 Observation results

The subjects took the perplexed behaviors when they lost the operational method. The representative cases were as the following. :

① The inputted Roman characters can't be transformed in the correct Chinese character.
② The double consonant in the Japanese syllabary can't be inputted with Roman characters.
③ The sentences can't be edited.
④ The operation mode once changed can't be returned to origin.

The representative perplexed behaviors were as the following. :

① Change of the facial expression without the key input : knit the eyebrows, open the eyes wide, open the mouth, smile, laugh while showing the teeth
② Motion of the upper half of the body without the key input : bend forward to gaze on the CRT display, bend backward
③ Motion of the head without the key input : shake, tilt
④ Change of the glance without the key input : gaze on the CRT display, gaze on the keyboard, move between the CRT display and the keyboard frequently
⑤ Motion of the arms without the key input : support the jaw, hold the head
⑥ Keyboard operation : take the longer time interval for key input, hit a same key repeatedly, hit keys at random

The words such as "Why?" are uttered in the perplex situations. However, the voice signal is excluded here because it is easily disturbed by other person's voice and noise in the practical scene.

Table 1 shows the number of the subjects who take the above behaviors. Every subject took a few kinds of the above perplexed behaviors. The perplexed behaviors other than the item ⑥ "Keyboard operation" are attended with the break of the key input for more than 3 sec. While the jaw or the head is supported by the arm, the head position and attitude keeps still. Accordingly, supporting the jaw and holding the head can be classified as motion of the glance "gaze". It is clear that the detection of the perplex situations is possible 100% by combining the motion of the head and the time interval of the key input, while the rate of the subjects who don't change the facial expression is 29% (9 persons).

**Table 1 Observation results of perplexed behaviors.**

| object and behavior | | | number of persons (overlapped) | | |
|---|---|---|---|---|---|
| motion | upper half body | bend backward | 8 | 31 | 31 |
| | | bend forward | 13 | | |
| | glance | come and go | 7 | | |
| | | gaze | 26 | | |
| | face | tilt | 15 | | |
| | | shake | 2 | | |
| | key stroke | same key | 12 | 14 | |
| | | random key | 2 | | |
| facial expression | knit the eyebrows | | 4 | 22 | |
| | open the eyes wide | | 1 | | |
| | open the mouth | | 12 | | |
| | smile | | 6 | | |
| | laugh while showing the teeth | | 8 | | |

## 3. RECOGNITION OF HEAD MOTIONS

### 3.1 Chase of both pupils

According to our observation, it became clear that the state of mind was shown in action patterns rather than the facial expression. We adopted the method of presuming the operator's perplex situations from his action. In our method, the movement of the head is captured by the image processing.

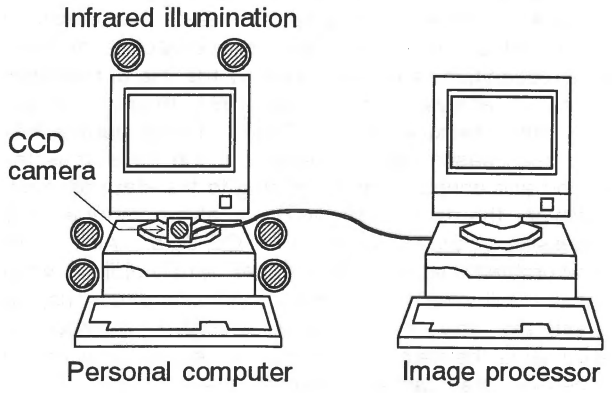



**Figure 2 Experimental system for detecting perplexed behaviors.**

In general, it is difficult to extract the operator's head by image processing in the practical scene. We captured the movement of both pupils instead of the head motion, because the pupils move together with the head and can be easily extracted. Figure 2 illustrates the setup for the recognition of the head motions. A monochrome CCD camera (XC-75 : Sony Co.) is set between the CRT display and the keyboard of the personal computer for the word processor. The camera is turned upward and takes the operator's head. Even in the case that the operator's face turns toward the keyboard, both pupils can be taken by such the camera position and attitude. To extract the pupils in distinction from the irises, the images are taken in near infrared band. The high-pass (visible transmitting, infrared absorbing) filter of the camera is demounted and a low-pass (infrared transmitting, visible absorbing) filter (IR-85: HOYA Co.) is attached to the camera. IR-LEDs (infrared light emitting diodes: AN304: Stanley Co.) of which the spectral emissivity peaks at a wavelength of about 950 nm are used as light sources. The light sources are attached around the personal computer. Image processing system consists of an image processing board (GPB-1 : Sharp Semiconductor Co.) with an image processing library, a personal computer (DESKPRO 4/33i : COMPAQ Co.) and a C language compiler (MS-C/C++: Microsoft Co.). The ranges of change in the size and the form of both pupils, the mutual distance and the inclination of the straight line which ties both pupils were measured for all sorts of the attitude of some persons. The ranges with some margins are defined as the conditions for extracting both pupils.

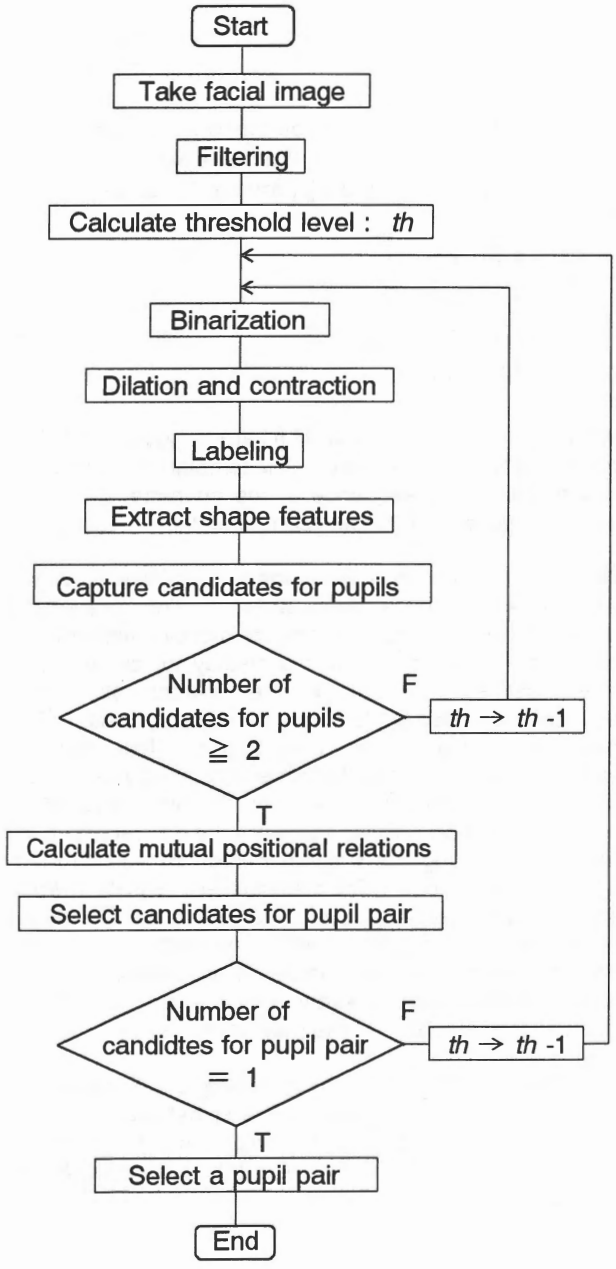



**Figure 3 Flow chart for extracting both pupils.**

Figure 3 shows the flow chart for extracting both pupils. First of all, binarization of the input image is done at the threshold level based on the gray levels of the 50 sampling points in the image. Next, unevenness of objects and the number of holes and breaks in the objects are reduced by doing dilation and contraction. Then, labeling is done to the objects in the image. Candidates for pupils are captured on the basis of their shape features: their peripheral length, their circleness. If the number of the candidates for pupils are 2 or more, then a suitable pair for both pupils is extracted on the basis of the mutual positional relations: Euclid distance between the objects, the inclination of the straight line which ties the objects. If the number of candidates for pupils are less than 2, or if the number of candidates for pupil pair isn't 1, the binarization is done again by subtracting 1 from the threshold level. In order to keep extracting both pupils in real time, the regions of interest on and after the second frame are limited around both pupils by the following method.: Let $(r_x, r_y)$ and $(l_x, l_y)$ be the coordinates of the right pupil and the left pupil, respectively (see Figure 4). Then the rectangular region which is defined by the following inequalities is selected as the region of interest on next frame.

$$r_x - \frac{l_x - r_x}{2} \le x \le l_x + \frac{l_x - r_x}{2} \quad (1)$$

$$\min(r_y, l_y) - \frac{l_x - r_x}{2} \le y \le \max(r_y, l_y) + \frac{l_x - r_x}{2} \quad (2)$$

where $l_x > r_x$ .

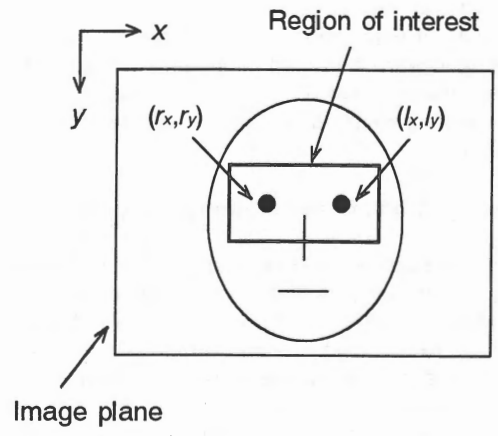
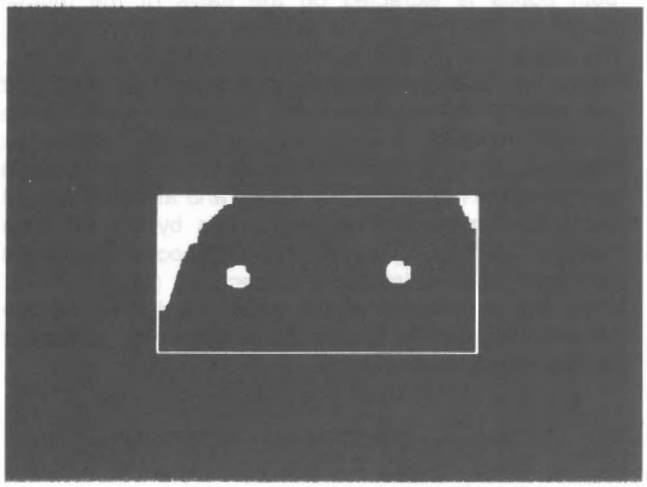



**Figure 4 The region of interest.**

Figure 5 (a) and (b) show the labeled objects in the whole binary image and in the clipped binary image around both pupils.

**(a) in the whole image**



**(b) in the clipped image around both pupils**

**Figure 5 Objects in binary image.**

There are only a few objects in the clipped image, so that the processing speed is enhanced. As the result of experiments, it was made clear that both pupils moving with the head were extracted continually without markers. The time interval between the frames is about 0.09 sec. This processing speed is fast enough to chase both pupils in real time.

### 3.2 Recognition of head motions using DP matching

The motions such as shaking, tilting, bending backward, bending forward to gaze on the CRT display, keeping still and nodding are taken as the basic actions in the perplex situations. Nodding shows the motion of head with the change of the glance between the CRT display and the keyboard frequently. To capture these 6 motions from the image sequence, the displacement velocity in 4 directions such as $x$ (right and left), $y$ (up and down), $z$ (forth and back), and $\theta$ (the rotation about $z$ axis) are calculated. The positions in $x$ and $y$ direction are obtained by calculating the coordinates of the middle point $(x_0, y_0)$ between both pupils.

$$x_0 = \frac{r_x + l_x}{2} \tag{3}$$

$$y_0 = \frac{r_y + l_y}{2} \tag{4}$$

Since the distance between both pupils corresponds to the distance between the face and the camera, we treated Euclid distance between both pupils as the relative position $z_0$ in $z$ direction. The rotation angle $\theta_0$ is calculated by using the inclination of the straight line tied between both pupils.

$$z_0 = \sqrt{(l_x - r_x)^2 + (l_y - r_y)^2} \tag{5}$$

$$\theta_0 = \arctan \frac{l_y - r_y}{l_x - r_x} \tag{6}$$

Let $v_x[i]$, $v_y[i]$, $v_z[i]$, and $v_t[i]$ be the displacement velocity in $x$, $y$, $z$, and $\theta$ direction in the frame No.$i$, respectively. Then, $v_x[i]$, $v_y[i]$, $v_z[i]$ and $v_t[i]$ are expressed as

$$v_x[i] = x_0[i] - x_0[i-1] \tag{7}$$

$$v_y[i] = y_0[i] - y_0[i-1] \tag{8}$$

$$v_z[i] = z_0[i] - z_0[i-1] \tag{9}$$

$$v_t[i] = \theta_0[i] - \theta_0[i-1] \tag{10}$$

where $x_0[i]$, $y_0[i]$, $z_0[i]$, and $\theta_0[i]$ are $x_0$, $y_0$, $z_0$, and $\theta_0$ in the frame No.$i$, respectively. The template of each action is defined as the sequence of the characteristic vector which is made up of these 4 components.

The template of each action is made as follows. : One of the subjects repeats each action. The relationship between the frame number and the displacement velocity in 4 directions is shown in the display for every action. The most suitable part in the characteristic vector sequence is picked out for every behavior by hand. Each basic action takes from 1 to 2 sec. The length of templates is determined 20 frames in consideration of the time interval between frames and the time required for each basic action. Since there are the differences in the magnitude among the 4 components, the normalization has to be carried out. The displacement velocity of which the absolute value is largest of all action pattern data is picked up for every component. The coefficient of which the maximum absolute value is converted to 1 is multiplied to all data for every component. Figure 6 (a)-(f) show the templates of the head motions.

DP (Dynamic Programming) matching is utilized here as the method of the pattern matching between unknown input vector sequence and the templates. The unknown input pattern $\boldsymbol{T}$ and one of the template patterns, $\boldsymbol{R}$, are expressed as

$$\boldsymbol{T} = \boldsymbol{a}_1, \boldsymbol{a}_2, \boldsymbol{a}_3, \cdots, \boldsymbol{a}_i, \cdots, \boldsymbol{a}_I$$

$$\boldsymbol{R} = \boldsymbol{b}_1, \boldsymbol{b}_2, \boldsymbol{b}_3, \cdots, \boldsymbol{b}_j, \cdots, \boldsymbol{b}_J$$

where $a_i$ and $b_j$ are the characteristic vector of the frame No.*i* in pattern ***T*** and the characteristic vector of the frame No.*j* in pattern ***R***, respectively. This time both *I* and *J* are 20. Each time a new frame is taken, unknown input pattern ***T*** is updated. The vectors of unknown input pattern ***T*** are converted by multiplying the coefficients which are determined in the above normalization.

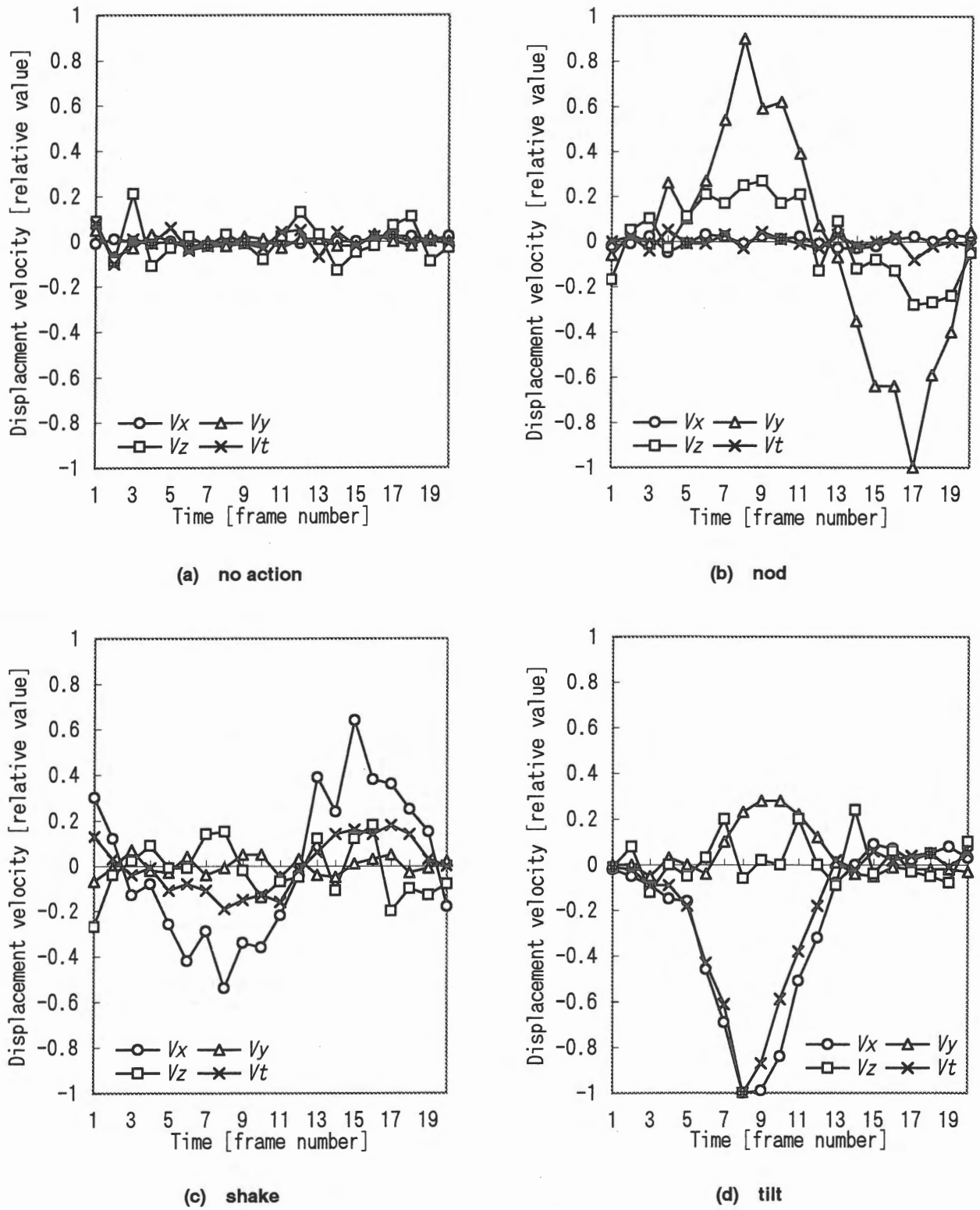
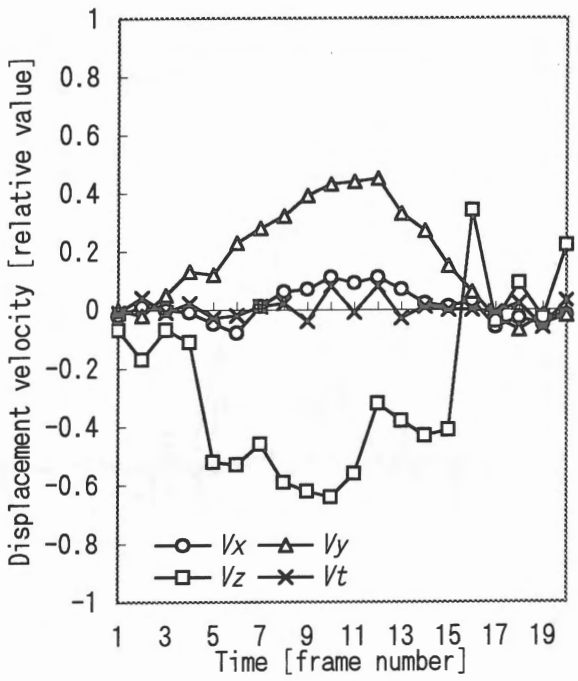
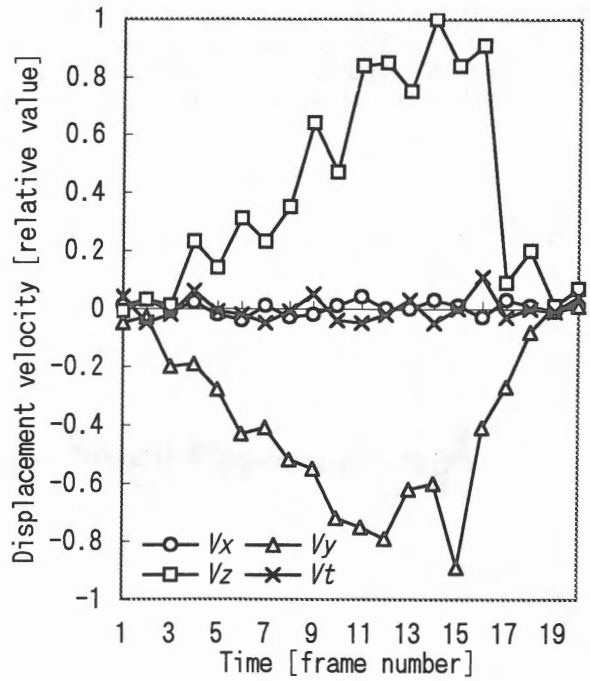



**Figure 6 Templates of head motions.**




(e) bend backward




(f) bend forward

**Figure 6 Templates of head motions.**

Let $d(i, j)$ be the distance between vector $a_i$ and vector $b_j$, then $d(i, j)$ is calculated by the following equation.

$$\begin{aligned} d(i,j) &= |a_i - b_j| \\ &= \{(v_{ax}[i]-v_{bx}[j])^2 + (v_{ay}[i]-v_{by}[j])^2 \\ &\quad + (v_{az}[i]-v_{bz}[j])^2 + (v_{at}[i]-v_{bt}[j])^2\}^{\frac{1}{2}} \end{aligned} \quad (11)$$

where

$$a_i = (v_{ax}[j], v_{ay}[j], v_{az}[j], v_{at}[j])$$
$$b_j = (v_{bx}[j], v_{by}[j], v_{bz}[j], v_{bt}[j])$$

Let $D(T, R)$ be the distance between pattern $T$ and pattern $R$, then $D(T, R)$ is calculated by the following equation.

$$D(T,R) = \frac{\min \sum_{i=1}^{I} d(i,j)}{I} \quad (12)$$

The template pattern which gives the least distance between patterns is judged as the input motion pattern.

The recognition experiment of the head motions were done for the specified person whose action patterns are adopted as the templates and 3 unspecified persons.

Table 2 shows the results of the matching distances between the unknown input patterns and templates in the case of the specified person. When the real actions were the same kind as the templates, the matching distances became shortest. Tabel 3-5 show the results of matching distance in the cases of unspecified persons. In these cases, the matching distances also became shortest between the same kind of action.

**Table 2 Distances between input action patterns of specified person and templates.**

| input \ template | keep still | nod | shake | tilt | bend backward | bend forward |
|---|---|---|---|---|---|---|
| keep still | **0.04** | 0.25 | 0.42 | 0.29 | 0.34 | 0.26 |
| nod | 0.36 | **0.22** | 0.53 | 0.48 | 0.50 | 0.44 |
| shake | 0.35 | 0.46 | **0.25** | 0.51 | 0.54 | 0.47 |
| tilt | 0.29 | 0.43 | 0.53 | **0.11** | 0.46 | 0.39 |
| bend backward | 0.41 | 0.54 | 0.64 | 0.53 | **0.16** | 0.63 |
| bend forward | 0.23 | 0.33 | 0.46 | 0.39 | 0.48 | **0.14** |

**Table 3 Distances between input action patterns of unspecified person A and templates.**

| input \ template | keep still | nod | shake | tilt | bend backward | bend forward |
|---|---|---|---|---|---|---|
| keep still | **0.05** | 0.24 | 0.40 | 0.26 | 0.34 | 0.23 |
| nod | 0.42 | **0.29** | 0.60 | 0.53 | 0.49 | 0.54 |
| shake | 0.36 | 0.59 | **0.26** | 0.61 | 0.71 | 0.61 |
| tilt | 0.35 | 0.47 | 0.61 | **0.15** | 0.50 | 0.47 |
| bend backward | 0.38 | 0.46 | 0.67 | 0.50 | **0.17** | 0.61 |
| bend forward | 0.26 | 0.32 | 0.44 | 0.39 | 0.52 | **0.15** |

**Table 4 Distances between input action patterns of unspecified person B and templates.**

| input \ template | keep still | nod | shake | tilt | bend backward | bend forward |
|---|---|---|---|---|---|---|
| keep still | **0.09** | 0.27 | 0.44 | 0.22 | 0.39 | 0.25 |
| nod | 0.43 | **0.33** | 0.58 | 0.61 | 0.48 | 0.55 |
| shake | 0.35 | 0.67 | **0.26** | 0.49 | 0.66 | 0.70 |
| tilt | 0.41 | 0.72 | 0.66 | **0.30** | 0.48 | 0.46 |
| bend backward | 0.43 | 0.53 | 0.44 | 0.41 | **0.27** | 0.56 |
| bend forward | 0.40 | 0.47 | 0.55 | 0.36 | 0.48 | **0.22** |

**Table 5 Distances between input action patterns of unspecified person C and templates.**

| input \ template | keep still | nod | shake | tilt | bend backward | bend forward |
|---|---|---|---|---|---|---|
| keep still | **0.07** | 0.21 | 0.39 | 0.31 | 0.36 | 0.24 |
| nod | 0.44 | **0.24** | 0.56 | 0.53 | 0.52 | 0.49 |
| shake | 0.38 | 0.54 | **0.28** | 0.52 | 0.57 | 0.59 |
| tilt | 0.37 | 0.52 | 0.51 | **0.20** | 0.44 | 0.35 |
| bend backward | 0.33 | 0.60 | 0.57 | 0.44 | **0.22** | 0.50 |
| bend forward | 0.30 | 0.42 | 0.62 | 0.48 | 0.46 | **0.19** |

As a result, it was made clear that each head motion was recognized by applying DP matching to the displacement velocity vectors of both pupils. The proposed method is available for not only the specified person but also unspecified persons. The average time required for total processing from the take of the image to the indication of the matching result is about 0.11 sec per frame. This processing speed makes it possible to recognize the head motions in real time.

## 4. CONCLUSION

We have proposed the method for recognizing the perplex situations in word processor work. By observing the subjects in word processor work, it was made clear that the perplexed behaviors were mainly shown in their head motions. To chase both pupils and set the region of interest around them make it possible to capture the head motion in real time. The distances between the unknown input motion pattern and the template patterns are calculated by DP matching. As the result of DP matching, it was made clear that the basic actions of perplexed behaviors were recognized. The proposed method is available for not only the specified person but also unspecified persons. To recognize the perplex situations correctly, it is desirable to combine the program which measures the time interval between the key strokes with this program for recognizing head motions. The method can be applied to the development of the software which responses automatically when the operator falls into the perplex situations.

In the case that the DP matching is implemented every frame, matching results are often incorrect in the beginning of action. It is desirable that the recognition of the head motions is done based on the matching result data for several frames such as the frequency or the number of times the same action is selected.

Most people keep still after they "tilt", "bend forward" and "bend backward". In contrast to these actions, "keeping still", "nodding" and "shaking" are continued for a while. For these reasons, the connection to "keeping still" and the continuity of the same action are thought to be significant. It is thought that the perplex situations can be recognized more surely by making the sequences of action names and executing the pattern matching of the sequences.

## REFERENCES


Choi, C., Harashima, H. and Takebe, T., 1991. Analysis of facial expressions using three-dimensional facial model. IEICE Transactions on Information and Systems (D-II), J74-D-II (6), pp.766-777.

Kamitani, T. and Marutani, Y., 1994. Detection of emotive change by capturing the blinking intervals. Record of the '94 Kansai-Section Joint Convention of IEE Japan, p.G283.

Kamitani, T. and Marutani, Y., 1995a. Discrimination of human intention using facial images. Proc. of the '95 IEICE General Conf., p.A-257.

Kamitani, T. and Marutani, Y., 1995b. Detectability of the annoyed state by video images. Proc. of the 39th Annual Conf. of ISCIE, pp.545-546.

Kamitani, T. and Marutani, Y., 1995c. Analysis of perplexed behavior by DP matching. Record of the '95 Kansai-Section Joint Convention of IEE Japan, p.G345.

Kamitani, T. and Marutani, Y., 1996. Analysis of perplex situations in word processor work using facial images. Image Labo, Vol.7, No.4, pp.324-334.

Kamitani, T. and Marutani, Y., 1997a. Recognition of basic actions for the detection of the personal difficulty using the position of pupils. Technical Report of IEICE, HCS96-41, pp.19-26.

Kamitani, T. and Marutani, Y., 1997b. Recognition of basic actions for the detection of the personal difficulty using the position of pupils. Proc. of the '97 IEICE General Conf., p.A-14-6.

Kamitani, T. and Marutani, Y., 1997c. Analysis of perplex situations in word processor work using facial image sequence. Proc. of SPIE: Human Vision and Electronic Imaging II (EI'97), Vol.3016, pp.324-334.

Kamitani, T. and Marutani, Y., 1997d. Recognition of perplexed behaviors using DP matching. Proc. of the 36th SICE Annual Conf., Domestic Session Papers Vol.1, pp.377-378.

Kamitani, T. and Marutani, Y., 1997e. Recognition of perplexed behaviors by DP matching, Human Interface News and Report, Vol.12, No.4, pp.475-482.

Kitamura, Y., Ohya, J. and Kishino, F., 1993. A study of grabbing facial actions from facial images with genetic programming. Technical Report of IEICE, PRU93-65, pp.23-28.

Matsuno, K., Lee, C. and Tsuji, S., 1993. Recognition of human facial expressions with potential net. Technical Report of IEICE, PRU93-64, pp.17-22.

Ochiai, M., Sugouchi, M., Haji, M. and Shimizu, E., 1994. Behavioral properties in perplex situation according to degrees of expertise of word processor —The analyses of perplexed behavior in the use of word processor in personal computer—. Fac. Lett. Rev. Otemon Gakuin Univ., Vol.30, pp.45-68.