FACIAL EXPRESSION RECOGNITION FROM IMAGE SEQUENCES USING SELF-ORGANIZING MAPS

Ayako KATOH*, Yasuhiro FUKUI**
*Student, **Professor, Applied Systems Engineering, Science and Engineering
Graduate School of Tokyo Denki University
Ishizaka, Hatoyama, Hiki, Saitama, 355-0311
E-mail: aya@bme.f.dendai.ac.jp, fukui@f.dendai.ac.jp
Japan

Commission V, Working Group V/4

KEYWORDS: facial expression, image sequences, self-organizing map, neural networks

ABSTRACT

Just as humans use body language or nonverbal language such as gestures and facial expression in communication, in the future computers will also be able to communicate with humans. In medical engineering, it is possible that recognition of facial expression can be applied to support communication with persons who have trouble communicating verbally such as infants, aged persons and mental patients. The purpose of this study is to enable recognition of human emotions by facial expressions using engineering methods. Our observations of facial expressions founds that the important facial segments for recognition are the eyebrows, the eyes and the mouth. We also found that it is important to recognize facial expressions by identifying changes in facial expressions by using sequences of images. Self-organizing maps, which are neural networks of two-layer structure, are used to extract features of image sequences. A self-organizing map is taught for each facial segment. The image sequences of six types of facial expressions are recorded on VTR and made into image sequences consisting of 30 images per second. Subjects are asked to make facial expressions ranging from expressionless to one of the six types of facial expressions under consideration. Gray levels of each segment are input into the self-organizing map corresponding to each segment. The neuron in the output layer, called the victory neuron, reacts to the feature nearest the input segment reacts. Our analysis of the changes in victory neurons demonstrates that they have characteristic features which correspond to each of the six facial expressions.

1. INTRODUCTION

Sensitivity information such as emotions plays an important role in communication. The means for conveying the sensitivity information are nonverbal languages such as gestures, facial expressions and voice tone. Until now, research on sensitivity information has been carried out primarily in the field of psychology (Hiroshi et al., Y., 1994).

But in recent years, it has also been taken up in the fields of engineering. Focus has been shifted from artificial intelligence based on knowledge provided by the researcher towards intelligence based on self-generated knowledge in computer science. In the field of medical engineering the mental condition of patients is considered in addition to his or her physical condition (Noboru, 1994). Recently extraction of sensitive information shown by nonverbal language is given considerable attention in the field of human interface (Shigeo, 1994).

The purpose of this study is to recognize human emotions in facial expressions by focusing on facial expressions as a form of nonverbal language.

Research on facial expressions has been carried out in the field of psychology. Most of the research discusses emotions manifested in facial expressions using a instant image (Hitoshi et al, 1994. Katsuhiro et al., 1994). The six types of facial expressions shown by Ekman et.al. are famous (Paul et al., 1975). Many studied have also been made on facial expression recognition using an instant image. But, a number of studied have pointed out limitations in recognizing facial expressions using only an instant image (Hiroshi, K. et al., 1995. Hiroshi, K. et al., 1993. Kenji, 1991. Tatsumi, 1995.). Thus, in this study, an attempt was made to recognize facial expressions from image sequences.

Our observations demonstrate the important of recognizing facial expressions from movements of several facial segments. A method for extraction features using self-organizing maps is reported this paper.

2. SELF-ORGANIZING MAP

Self-organizing maps are neural networks designed to input data into one or two dimensional space while maintaining the same distance found in the original dimensional space (Teuvo, 1996). Two-layer of self-organizing maps consists of input and output layer. All

neurons in the input layer are connected to all neurons in the output layer. For some input data, only one neuron reacts in the output layer. The neuron that reacts is the one which has the features nearest the features of the input data. This is called the victory neuron.
Self-organizing maps are learned by a competition method that has no teaching data. The characteristic feature of this method is that the positions of neurons in the output layer are related. The victory neuron and its neighborhood neurons are learned in a group. Self-organizing maps are unnecessary for obtaining information about categories of input data. Resemblance of input data is formed by itself in the output layer. The neurons placed in a neighborhood in the output layer have features nearest each other and the neurons placed away from each other in the output layer have different features. Since the category of the input data is not given while the self-organizing map is being learned, it is possible to design the self-organizing map so that it will reflect distribution conditions of input data. In other words, the character of a self-organizing map depends on its set of input data. The set of input data should therefore be examined closely.

3. METHODS

3.1 Outline

Figure 1 shows the outline of this method. Image sequences of changing facial expressions are prepared. Positions of facial segments such as eyebrows, eyes and mouth, are already known and rectangular segments which include each facial segment are selected. For example, an image of the rectangular segment of an eyebrow is input into a self-organizing map which is learned to classify features of the eyebrow. Then, a feature of the image is represented by locating in the output layer the self-organizing map of the victory neuron that we refer to as its number. These procedures are applied to other facial segments as well. By applying these procedures on all Image sequences, the changes in the feature corresponding to the changes in the facial expressions is considered to be the changes in the victory neuron number.
Figure 1 summarizes this method. First, the eyebrows, the eyes and the mouth of the facial images are input into the self-organizing map. Facial expressions are classified by analyzing the change of victory neuron number.

3.2 Learning Self-organizing Maps

Important facial segments for recognition are the eyebrows, the eyes and the mouth. The self-organizing map for each segment is learned.
In this study, input images for learning are image sequences that change from expressionless to one of the six types of facial expressions, and there are 30 such images per second. It is necessary to prepare various kinds of input image to give self-organizing maps an ability to classify many kinds of facial expressions, including those which express transitional periods.
Figure 2 shows the learning process of self-organizing maps. Initial conditions of self-organizing maps are given at random. A rectangular segment that includes an eyebrow is selected from the input image and is converted into a gray image consisting of 256 levels. Then intensity of each picture element of the selected segment is input into the self-organizing map of the eyebrow. The neuron in the output layer with the feature nearest the feature of the input image becomes the victory neuron. The features of the neurons in the output layer are defined by the interconnection weights of the input and output layers. The interconnection weights of the victory neuron and the neighborhood neurons are changed in order to make the features of these neurons approximate those of the input image by using Equation 1. w_i is an i th interconnection weight in the output layer. N_c is a set of numbers of which interconnection weights are changed. In the beginning of learning, size of the N_c is large and it becomes small as its learning makes progress. α is a coefficient of learning. The α becomes big as learning makes progress. A self-organizing map is formed by repeating this process changing input image.
The self-organizing maps which correspond to the eyes

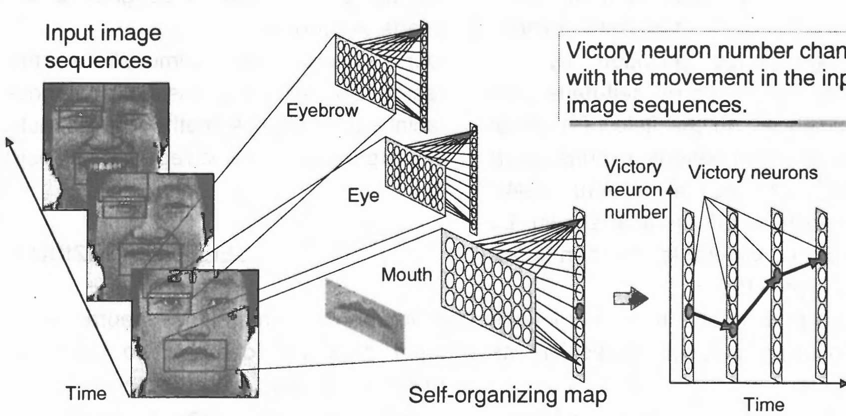


Figure 1 Feature Extraction Method

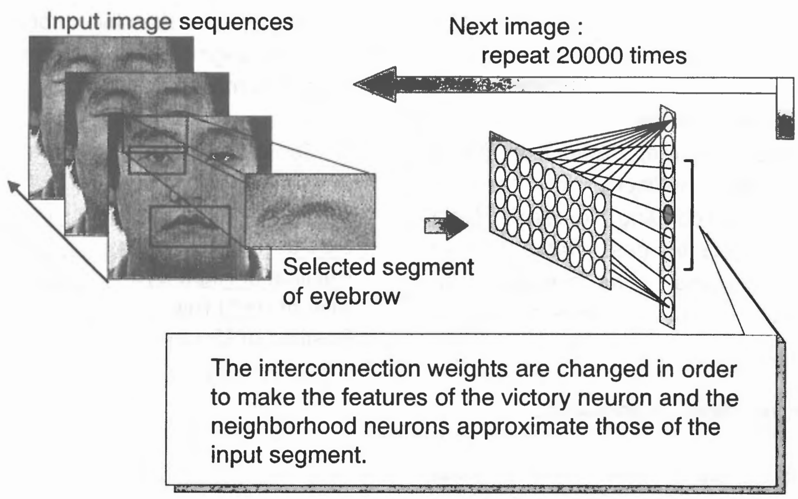


Figure 2 Learning Self-organizing Maps

$$w_i(t+1) = \begin{cases} \dfrac{w_i(t) + \alpha'(t)x(t)}{\left\| w_i(t) + \alpha'(t)x(t) \right\|} & if \; i \in N_c(t) \\ w_i(t) & if \; i \notin N_c(t) \end{cases} \quad \cdots 1$$

and the mouth are prepared through the same learning process.

3.3 Feature Extraction method

The features of the image sequences of recognition targets are extracted by using learning self-organizing maps. Positions of facial segments such as eyebrows, eyes and mouth in facial images are supposedly already known, and rectangular segments that include important facial segments are selected. The features of rectangular segments are expressed in victory neuron number by inputting into learning self-organizing maps the intensity of elements in rectangular segments. This process is carried out for all image sequences, and changes in the victory neuron number corresponding to those in the facial expression are analyzed. To recognize facial expressions, it is necessary to take the movements of all facial segments into consideration. Thus, the eyebrows and the eyes and the mouth of the neuron number constitute the axes and the feature of the facial expression in time is represented by a three-dimensional point which is defined by the victory neuron number of the eyebrow, the eye and the mouth. The point will move as facial expression change. Facial expressions are classified by analyzing the movement of the point.

4. DISCRIMINATION EXPERIMENT

4.1 Making Image Sequences of Facial Expressions

Subjects are ordinary people without formal training in acting. The filming is done indoors. Lights are placed to make the brightness in front of the face 700-lux. A video camera is placed directly in front of the subject's face. Subjects are asked to make expressions ranging from expressionless to one of the six types of facial expressions one at a time and while trying their best not to move their heads. Each facial expression is recorded three times, and emotions are expressed one at a time. Subjects are not asked to express multiple emotions or mixed emotions. Under these conditions, facial expressions are recorded on VTR. Video sequences last from two to four seconds. Video sequences are rendered as image sequences of 30 images per second. Figure 3 shows the subjects and

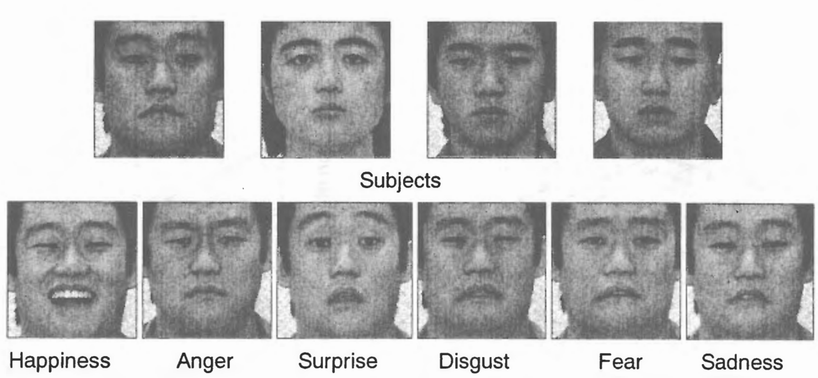


Figure 3 Subjects and Their Facial Expressions

their facial expressions.

4.2 Learning Self-organizing Maps

The learning time self-organizing maps require is 20000. Table 1 shows detailed conditions of learning. Rectangular segments corresponding to each facial segment are selected from facial images. Table 2 shows structure of self-organizing maps. The self-organizing maps are designed 1-dementional. It makes a following analysis simple.

4.3 Feature Extraction of Facial Expressions

Features are extracted by using self-organizing maps learned, in accordance with the conditions discussed above. Rectangular segments for each facial segment are selected from input images at the same positions as those used in a learning process. The size of the rectangular segments is the same as the size of those used in a learning process.

Table 1 Details of Images Used for Learning

Size of Facial Image	320×240 [pixel]
Position of Right Eyebrow	106 , 51 [pixel]
Size of Right Eyebrow	60× 35 [pixel]
Position of Right Eye	115 , 81 [pixel]
Size of Right Eye	45× 25 [pixel]
Position of Mouth	130 , 141 [pixel]
Size of Mouth	80× 40 [pixel]

Table 2 Structure of Self-organizing Map

	Size of Input Neuron	Size of Output Neuron
Map of Eyebrow	60×35 [pixel]	40×1 [pixel]
Map of Eye	45×25 [pixel]	40×1 [pixel]

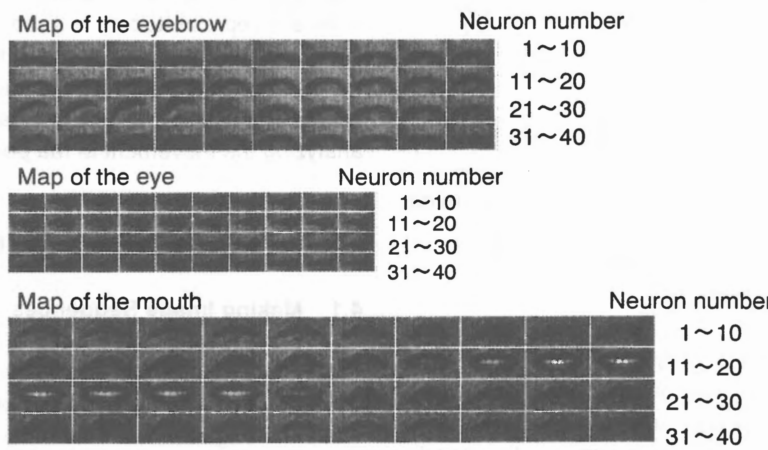
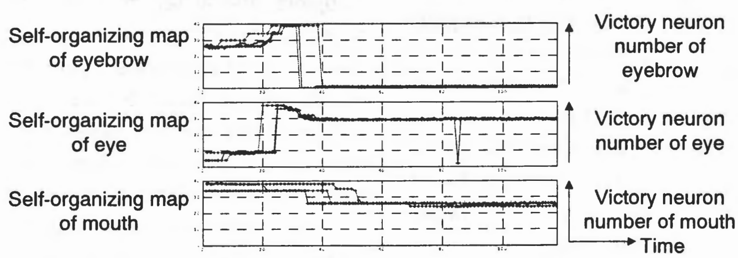
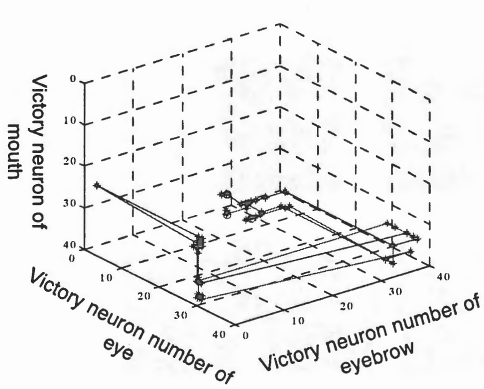


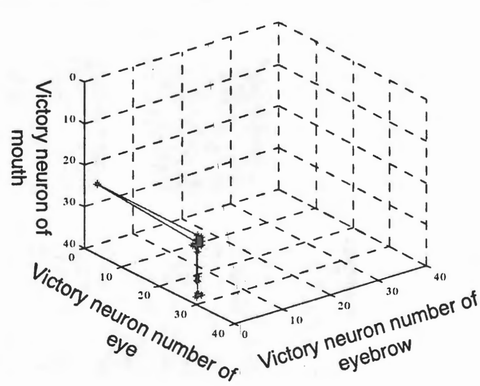
Figure 4 Subjects and Their Facial Expressions



(a) The movements of the victory neuron for each facial segment

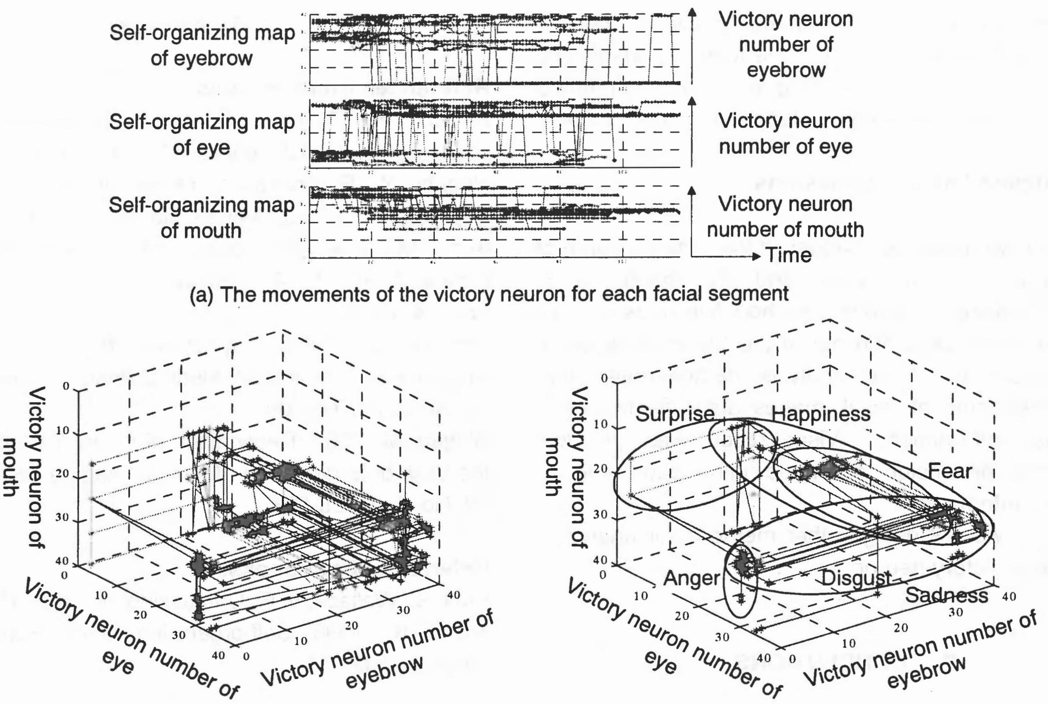


(b) The movements in 3-dementional space



(c) The movements when strongly expressed

Figure 5 The movements of the Victory Neuron(Anger)



(a) The movements of the victory neuron for each facial segment

(b) The movements when strongly expressed

(c) The movements when strongly expressed

Figure 6 The Classification of Facial Expressions

Map of Mouth	80×40 [pixel]	40×1 [pixel]

5. RESULTS

Figure 4 shows the self-organized maps used for recognition. Figure 5 shows the movement of victory neuron when the same kinds of facial expressions are input. The tracks of the victory neuron have similar shapes for three times. Figure (a) shows the movements of the each victory neuron. Figure (b) shows the movements of victory neuron in 3-dimensional space. Figure (c) shows the movements of the victory neuron while the facial expression was strongly expressed.

Figure 6 shows the movement of victory neuron when different kinds of facial expressions are input. The tracks of the victory neuron have features peculiar to each facial expression.

6. DISCUSSION

6.1 Validity of Recorded Facial Expressions

The facial expressions we recorded in this experiment ware based on the emotions we asked our subjects to express. To confirm that the facial expressions accurately expressed the emotions we've wanted, we sent questionnaire to 15 subjects. The latter ware instructed to watch a video image of each facial expression and facial expression they are watching. Table 3 summarizes our findings. Since four out of six types of emotions we interested (happiness, anger, surprise, fear) got high recognition rates, we concluded that emotions except sadness and fear ware effectively expressed. The results of our classification using self-organizing maps suggest that sadness and disgust are placed in the same category. The facial expressions of sadness ware classified by nether the self-organizing map or our human subjects. This suggests that sadness is a difficult emotion to express, and that facial expressions of sadness are not very pronounced. To test these hypotheses, we will need to develop better ways to record facial expressions.

Table 3 The results of questionnaires

The Kinds of Facial expression	Recognition rate of facial expressions[%]				
	Subjects				Average
	A	B	C	D	
Happiness	100	100	100	60	90
Anger	47	47	53	27	43
Surprise	93	80	87	60	80
Disgust	67	80	47	67	65
Fear	20	33	20	0	18
Sadness	13	20	27	13	18

6.2 The Learning Ability of Self-organizing Maps

The possibility of classifying facial expressions depends on how well self-organizing maps are learned. As shown in Figure 3, in some cases neurons placed far from in the output layer have similar features. When features are extracted using self-organizing maps, although image sequences are hardly changing, there are some cases where the movements of the victory neurons are so large that image sequences are perceived to be changing. Thus it is difficult to estimate changes image sequences from changes in the victory neuron. This problem might be overcome by considering learning conditions such as size of the self-organizing maps, the learning coefficient, and kinds of images using learning and changing method of weights. These we will be undertake in future studies.

6.3 Classifying Facial Expressions

In this study, we analyzed changes in the victory neuron of the eyebrows and the eyes and the mouth in 3-dimensional space. Using this method, five kinds of facial expressions were classified out of the basic six types of facial expressions. Our analysis demonstrates that important elements of facial images are reflected in 3-dimensional information. Visualization was possible reducing the dimension of image information into 3-dimensional information.
In the future, we will design other methods for analyzing the change of victory neuron.

7. CONCLUTIONS

In this study, we attempted to classify facial expressions by using self-organizing maps. By inputting various images of facial expressions into self-organizing maps and changing the interconnection weights, we were able to make self-organizing maps capable of classifying image features. This study demonstrates that by inputting images into self-organizing maps, it is possible to representing image features as a victory neuron number in self-organizing maps. Thus it is possible to consider changes in images as changes in the victory neuron number.
Movements of facial segments have features peculiar to each facial expression. The victory neuron showed changes peculiar to each facial expression when image sequences of facial expressions were input into self-organizing maps.
By analyzing the changes in the victory neuron, we were able to classify facial expressions, thus demonstrating the possibility of recognizing facial expressions by using this method. In the future, we will strive to develop an algorithm for recognizing facial expression and a method for automatically tracking such facial segments as the eyebrows and the eyes and the mouth.

References

References from Journals:
Hiroshi, K., Fumio, H., 1995. Monitoring of Facial Expressions. J.SICE, Vol.v34, No. 4, pp248-254.
Hiroshi, Y., Psychological Model of Recognizing Facial Expressions. J.SICE, Vol.33, No. 12, pp-1063-1069.
Kenji, M., 1991. Recognition of Facial Expression from Optical Flow. IEICE Transactions, Vol. E-74, No. 10, pp3474-3483.
Noboru, S., 1994. Expectation by a Psychiatrist to Recognition Face Image. Medical Imaging Technology, Vol. 12, No. 6, pp.700-709.
Shigeo, M., 1994. Recognition of Facial Expression –From the View of Engineering-. Medical Imaging Technology, Vol. 12, No. 6, pp688-693.

References from Books:
Paul, E., Wallace, V. F., Unmasking the Face. Printice-Hall.
Teuvo, K., 1996. Self-organizing Maps. Springer-Verlag, Tokyo, pp102-171.

References from Other Literature:
Hitoshi, O., Changsuk, C., Haruyuki, M., 1994. Computer Recognition of Facial Expressions. Conference on Imaging Engineering, pp23-26.
Hiroshi, K., Fumio, H., Susumu, I., 1993. Dynamic Recognition of 6 Basic Facial Expressions by Recurrent Neural Network. Technical Report of IEICE, HC92-59, pp11-16.
Katsuhiro, M., Chil-Woo, L., Saburo, T., 1994. Lecture Notes in Computer Science, Vol.800, Computer Vision – ECCV'94, pp513-520.
Tatsumi, S., 1995. Image Feature Extraction Using Wavelet Transformation and Its Application for Facial Expression Recognition. Technical Report of IEICE. IE94-147, pp15-22.