Orientation Selectivity for Representing Dynamic Diversity of Facial Expressions

H. Madokoro and K. Sato
Department of Machine Intelligence and Systems Engineering, Faculty of Systems Science and Technology, Akita Prefectural University, Yurihonjo, Japan
Email: {madokoro, ksato}@akita-pu.ac.jp

***Abstract*— This paper presents a representation method of facial expression changes using Adaptive Resonance Theory (ART) networks. Our method extracts orientation selectivity of Gabor wavelets on ART networks, which are unsupervised and self-organizing neural networks that contain a stability-plasticity tradeoff. The classification ability of ART is controlled by a parameter called the attentional vigilance parameter. However, the networks often produce redundant categories. The proposed method produces suitable vigilance parameters according to classification granularity using orientation selectivity. Moreover, the method can represent the appearance and disappearance of facial expression changes to detect dynamic, local, and topological feature changes from obtained whole facial images.**

***Index Terms*— Orientation Selectivity, Adaptive Resonance Theory, Gabor wavelets.**

I. Introduction

People with rich facial expressions are robust to uncertain situations or adverse circumstances. The roles of facial expressions in communication among people are important and various. Especially in a close relationship, we can mutually understand the feeling and intensity from the information of facial expressions. In the field of human communication, computer recognition of facial expressions has been studied for realizing a natural and flexible Man-Machine Interface (MMI) that can interpret the feeling or intensity of users [1].

Akamatsu defined facial diversity of two types [2]. Facial components such as eyes, eyebrows, and the mouth are different for each person. Facial features of those facial components' position, size, location, etc. are also different. This is called static diversity. On the other hand, we move facial muscles to express internal emotions unconsciously or express emotions as a message. Facial expressions are produced by the facial components and their transition from a normal facial expression. This is called dynamic diversity. Regarding facial recognitions in the field of facial image processing, only the use of static diversity is sufficient to obtain good results. For facial expression recognition, it requires not only static diversity but also dynamic diversity as a time-series to cope with facial pattern transitions.

This paper is based on "Orientation Selectivity for Representation of Facial Expression Changes," by H. Madokoro and K. Sato, which appeared in the Proceedings of the 2007 IEEE International Joint Conference on Neural Networks (IJCNN 2007), Orlando, Florida, USA, Aug. 2010. © 2007 IEEE.

Nishiyama et al. [3] proposed facial scores, a method to describe facial expression rhythms. They pointed out that facial expressions that are describable with a Facial Action Coding System (FACS) by Ekman [4] are only static features. Therefore, they did not use Action Units (AUs) of FACS. They originally used setting feature points because FACS can not describe time-series transitions of facial expressions. On the other hand, humans can recognize facial expressions to detect movements of local facial components from entire structures of faces. We do not need to detect facial elements as movements of characteristic points. We can automatically detect dynamic, local, and topological feature changes of facial expressions from whole facial changes.

Ekman defined six basic expressions (anger, sadness, disgust, happiness, surprise, and fear) based on basic feelings of six types [4]. However, the number of categories to express is unknown because facial expressions exist that are invalid or which reflect several mixed feelings. In this paper, we introduce Adaptive Resonance Theory (ART) networks [5] as a method to represent detection of dynamic, local, and topological changes of facial expressions. The ART, which was proposed by Grossberg et al., is a theoretical model of an unsupervised and self-organizing neural network to form a category adaptively in real time while maintaining stability and plasticity. Using incremental learning of ART, the method can classify facial expressions without presetting of the number of categories. In addition, facial expressions that are controlled by feelings change over time through aging. We consider that ART, which can learn over time, is useful to deal with time-series movements of facial expressions.

However, setting the parameters of ART networks is complex; furthermore, classification results depend strongly on settings and combinations of parameters. Especially, a parameter called the attentional vigilance parameter strongly influences classification granularity. In addition, ART networks generate redundant categories, even though the setting of vigilance parameters is the same. In this paper, we specifically describe orientation selectivity of Gabor wavelets for analyzing classification granularity of ART networks. The method can detect dynamic, local, and topological changes of facial expressions for category changes of ART networks. Moreover, the method can prevent redundant categories through the use of orientation selectivity.

doi:10.4304/jcp.7.9.2107-2113

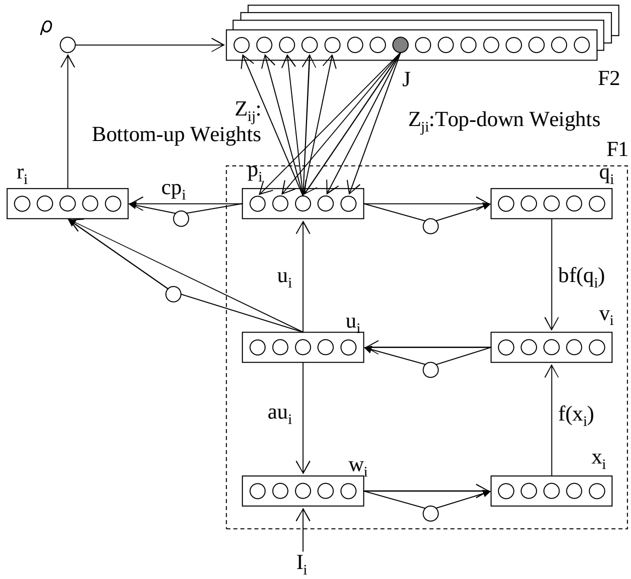
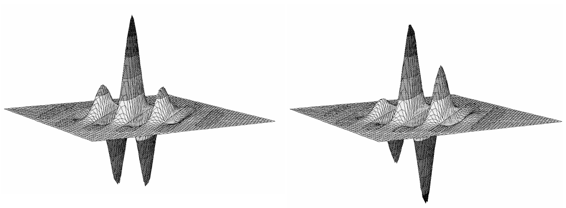


Figure 1. Architecture of an ART2 network.



(a) Real Part (b) Imaginary Part

Figure 2. Three-dimensional representations of Gabor wavelet filters.

II. Adaptive Resonance Theory 2

Actually, ART has many variations: ART1, ART1.5, ART2, ART2-A, ART3, ARTMAP, Fuzzy ART, Fuzzy ARTMAP, etc. [6]. We use ART2 [5], into which analog values can be input.

Figure 1 shows the ART network architecture. The network consists of two fields: Field 1 (F1) for feature representation and Field 2 (F2) for category representation. The F1 consist of six sub-layers: p_i, q_i, u_i, v_i, w_i, and x_i. These sub-layers realize Short Term Memory (STM), which enhances features of input data and detects noise for a filter. The F2 realizes Long Term Memory (LTM) based on finer or coarser recognition categories. The algorithm of ART2 is the following.

1) The top-down weights Z_{ji} and bottom-up weights Z_{ij} are initialized as

$$Z_{ji}(0) = 0, \qquad Z_{ij}(0) = \frac{1}{(1-d)\sqrt{M}}. \tag{1}$$

2) The sub-layers of F1 are initialized as

$$p_i(t) = q_i(t) = u_i(t) = v_i(t) = w_i(t) = x_i(t) = 0. \tag{2}$$

3) The input data I_i are presented to the F1. The sub-layers are propagated as

$$w_i(t) = I_i(t) + au_i(t-1), \tag{3}$$

$$x_i(t) = \frac{w_i(t)}{e + ||w||}, \tag{4}$$

$$v_i(t) = f(x_i(t)) + bf(q_i(t-1)), \tag{5}$$

$$u_i(t) = \frac{v_i(t)}{e + ||v||}, \tag{6}$$

$$q_i(t) = \frac{p_i(t)}{e + ||p||}, \tag{7}$$

$$p_i(t) = \begin{cases} u_i(t) & (inactive) \\ u_i(t) + dZ_{Ji}(t) & (active), \end{cases} \tag{8}$$

where

$$f(x) = \begin{cases} 0 & if \quad 0 \le x < \theta, \\ x & if \quad x \ge \theta. \end{cases} \tag{9}$$

4) Search for the maximum active unit T_J as

$$T_j(t) = \sum_j p_i(t) Z_{ij}(t), \tag{10}$$

$$T_J(t) = max(T_j(t)). \tag{11}$$

5) The weights Z_{ji} and Z_{ij} are updated as follows.

$$\frac{d}{dt} Z_{Ji}(t) = d[p_i(t) - Z_{Ji}(t)] \tag{12}$$

$$\frac{d}{dt} Z_{iJ}(t) = d[p_i(t) - Z_{iJ}(t)] \tag{13}$$

6) The output value of $r_i(t)$ is calculated as

$$r_i(t) = \frac{u_i(t) + cp_i(t)}{e + ||u|| + ||cp||}. \tag{14}$$

The reset is defined as

$$\frac{\rho}{e + ||r||} > 1. \tag{15}$$

7) If eq. (15) is true, the active unit is reset; go back to 4) to search again. If no active unit exists, a new category is created; return to 3). If eq. (15) is not true, repeat 3) and 5) until the changing of F1 is sufficiently small, then return to 2).

Parameters are the following: a and b are coefficients on feedback loops from u_i to w_i and from q_i to v_i; c is a coefficient from p_i to r_i; d is a learning rate; $cd/(1-d) \le 1$ is the constraint between them; and θ is a parameter to control a noise-detection level in layer v.

III. Gabor wavelets

Visual information captured by the retina is conveyed to Visual area 1 (V1) in the occipital lobe via the Lateral Geniculate Nucleus (LGN). The V1 consists of two visual cells: simple cells and complex cells. The LGN and simple cells have receptive fields. Receptive fields respond to a particular stimulus of figures such as the size, length, direction, movement direction, color, and frequency. This is called response selectivity. Since the time Hubel and Wiesel [8] discovered orientation selectivity on receptive fields from their electrophysiological experiment using anesthetized cats, orientation selectivity has become the most well known among response selectivity.

Various methods based on visual cortex information processing models have been proposed to develop image processing or computer vision systems [2], [7], [9]. The representation of Gabor wavelets, which can emphasize an arbitrary characteristic with inner parameters, is closed to receptive fields. Therefore, Gabor wavelets are applied to various fields such as character recognition, texture classification, and facial image processing [14], [15]. Gabor wavelets are functions that are combined with a plane wave propagating to one direction and a Gaussian wave. A three-dimensional (3D) representation of Gabor wavelets is shown in Fig. 2.

Let λ be a wavelength, and let σ_x and σ_y respectively denote widths of horizontal and vertical directions of Gaussian windows, where θ is the angle between the direction of a plane wave and the horizontal axis. The output of Gabor wavelets $G(x, y)$ is given as

$$G(x, y) = \exp\{-\frac{1}{2}(\frac{R_x^2}{\sigma_x^2} + \frac{R_y^2}{\sigma_y^2})\} \exp(i\frac{2\pi R_x}{\lambda}), \quad (16)$$

where

$$\begin{bmatrix} R_x \\ R_y \end{bmatrix} = \begin{bmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix}. \quad (17)$$

When Euler's formula,

$$\exp(i\theta) = \cos\theta + i\sin\theta, \quad (18)$$

is applied, the formula (16) is changed as:

$$G(x, y) = R_m(x, y) + iI_m(x, y), \quad (19)$$

$$R_m(x, y) = \exp\{-\frac{1}{2}(\frac{R_x^2}{\sigma_x^2} + \frac{R_y^2}{\sigma_y^2})\} cos(\frac{2\pi R_x}{\lambda}), \quad (20)$$

$$I_m(x, y) = \exp\{-\frac{1}{2}(\frac{R_x^2}{\sigma_x^2} + \frac{R_y^2}{\sigma_y^2})\} sin(\frac{2\pi R_x}{\lambda}). \quad (21)$$

The final output is

$$G(x, y) = \sqrt{Rm^2(x, y) + Im^2(x, y)}. \quad (22)$$

The suitable values of σ_x, σ_y are reported as a function of [11], so that

$$\begin{bmatrix} \sigma_x \\ \sigma_y \end{bmatrix} = \lambda \begin{bmatrix} S_x \\ S_y \end{bmatrix}, \quad (23)$$

where S_x and S_y are coefficients.

IV. Experiment

The purpose of this experiment is to detect facial expression changes for the category changes of ART networks from datasets that include both expressive and normal faces. Moreover, we evaluate orientation selectivity obtained from categorical changes of ART.

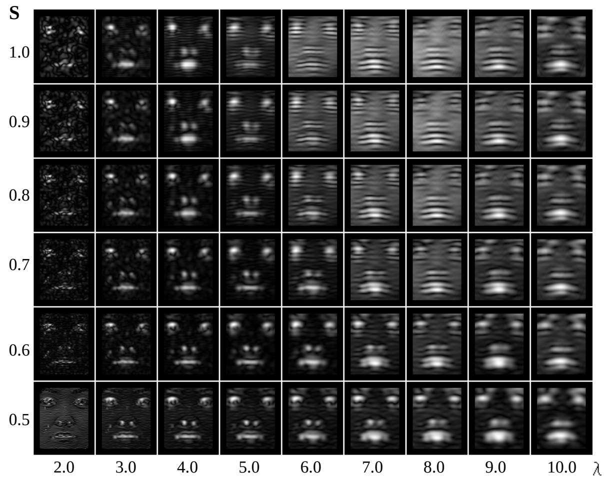


Figure 3. Gabor wavelet output images of the combination of λ and S.

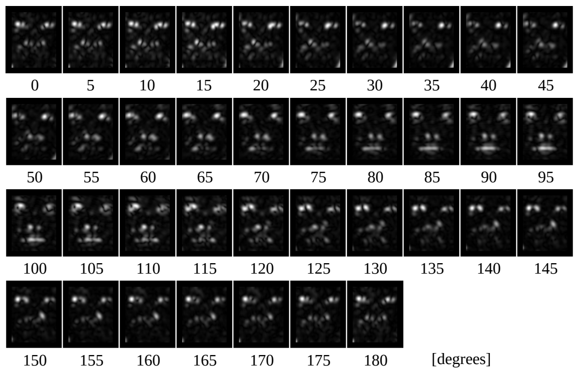


Figure 4. Gabor wavelet output images of θ ($0 \leq \theta \leq 180$, λ = 4.0 and S = 0.7)

A. Target Images

For this experiment, our evaluation targets are six basic facial expressions (anger, sadness, disgust, happiness, surprise, and fear) as defined by Ekman. We took 600 facial images from each person. The frame rate was 10 frames per second. Each facial expression comprises 100 images. The images at W320 $\times$ H240 pixels resolution were taken using a CCD camera in front of the face. We manually clipped the facial region of W92 $\times$ H110 pixels from the images. For automatic facial detection, we plan to use a method using Haar-like features by Papageorgiou et al. [12].

The targeted person is a woman in her 20s, a university

TABLE I.
TARGET FRAMES THAT PORTRAY FACIAL EXPRESSIONS.

Facial expressions	1st	2nd	3rd
Anger	18-30	50-57	76-82
Sadness	11-24	40-48	65-78
Disgust	15-32	52-65	90-100
Happiness	21-47	64-78	-
Surprise	16-26	52-60	81-92
Fear	16-34	56-63	85-98

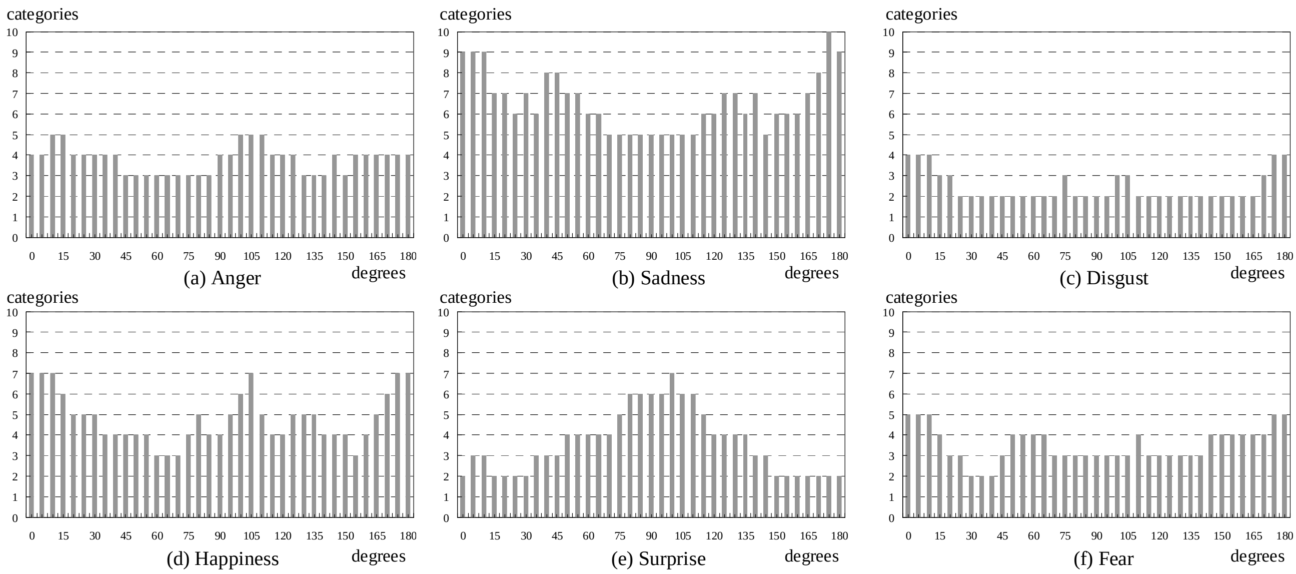


Figure 5. The number of categories in each direction from 0 to 180 degrees by 5 degree steps (ρ = 0.970).

graduate school student. She repeated one expressed face and a normal face in each facial expression. Therefore, this image dataset consists of two face types in each facial expression: one type of expression face and a normal face. The facial expressions are intentional. However, the timing of expression is idiosyncratic: the targeted person decided that timing. After the dataset acquisition, we specified appearance and disappearance points of all facial expression datasets. The appearance and disappearance points are summarized in Table I.

B. Parameters

We evaluated parameters of Gabor wavelets and ART2 networks. Figure 3 shows the relationship between λ and S in the case of $\theta = 0$. The respective ranges of λ and S are $2.0 \leq \lambda \leq 10.0$ and $0.5 \leq S \leq 1.0$. We selected λ = 4.0 and S = 0.7 (σ = 2.8) because these representations are sparse features. In this case, we set the parameters subjectively. Optimizing parameters using automatic and objective setting methods is a subject for our future work [13].

The electrophysiological knowledge indicates that the visual range of receptive fields is 1–5 degrees to yield a response to an input stimulation, the parameter θ is set in each case to five degrees. Figure 4 shows a two-dimensional representation of Gabor wavelets from 0 to 180 degrees by 5 degree steps. Moreover, we set the parameters of ART2 networks, θ = 0.01, a = b = 10, c = 0.225, d = 0.8, e = 0.0001, based on our experience and the Grossberg's original paper [5].

C. Results and Discussion

Figure 5 shows the number of categories in each direction from 0 to 180 degrees step by 5 degree steps. The vigilance parameter ρ is set to 0.970. The directions with a large number of categories mean that facial feature changes are large in the input images. The number of categories of surprise, which features the open level of a mouth at 90 degrees and nearby, is larger than for the other directions (Fig. 5e). This result means that the category changes of these directions are remarkable. The number of categories of sadness, which features wrinkles between the eyebrows at 0 and 180 degrees and nearby, is larger than for the other directions (Fig. 5b). The result of sadness is a large number of categories compared with other facial expressions. This result means that category changes of these directions are also remarkable.

Figure 6 shows category changes in the case of ρ = 0.970. This figure shows the generation of new categories as filled rectangles and transitions to existing categories as empty rectangles. The vertical lines in each graph are appearance or disappearance of facial expressions as specified in the section IV-A. These lines that correspond to Table I show the changing frames of appearance and disappearance between the normal expression and each facial expression. Category changes are not required on these lines because the appearance and disappearance continue a few frames before and after specified frames.

The appearance of anger is represented for categorical changes of ART networks in all three times (Fig. 6a). The directions at the first appearance are only three: 5, 10, and 15 degrees. The directions at the second and third appearance include wide ranges of directions, meaning that the ranges of orientation selectivity are narrow at the first point and wide at the second and third points. The first and second disappearance of anger can be detected, but the third one can not be detected. The appearance and disappearance of sadness are represented (Fig. 6b). However, the setting value of ρ is high because many categories occurred, except at the transition points. The expression of disgust shows a weak response (Fig. 6c). This response means the classification granularity is low,

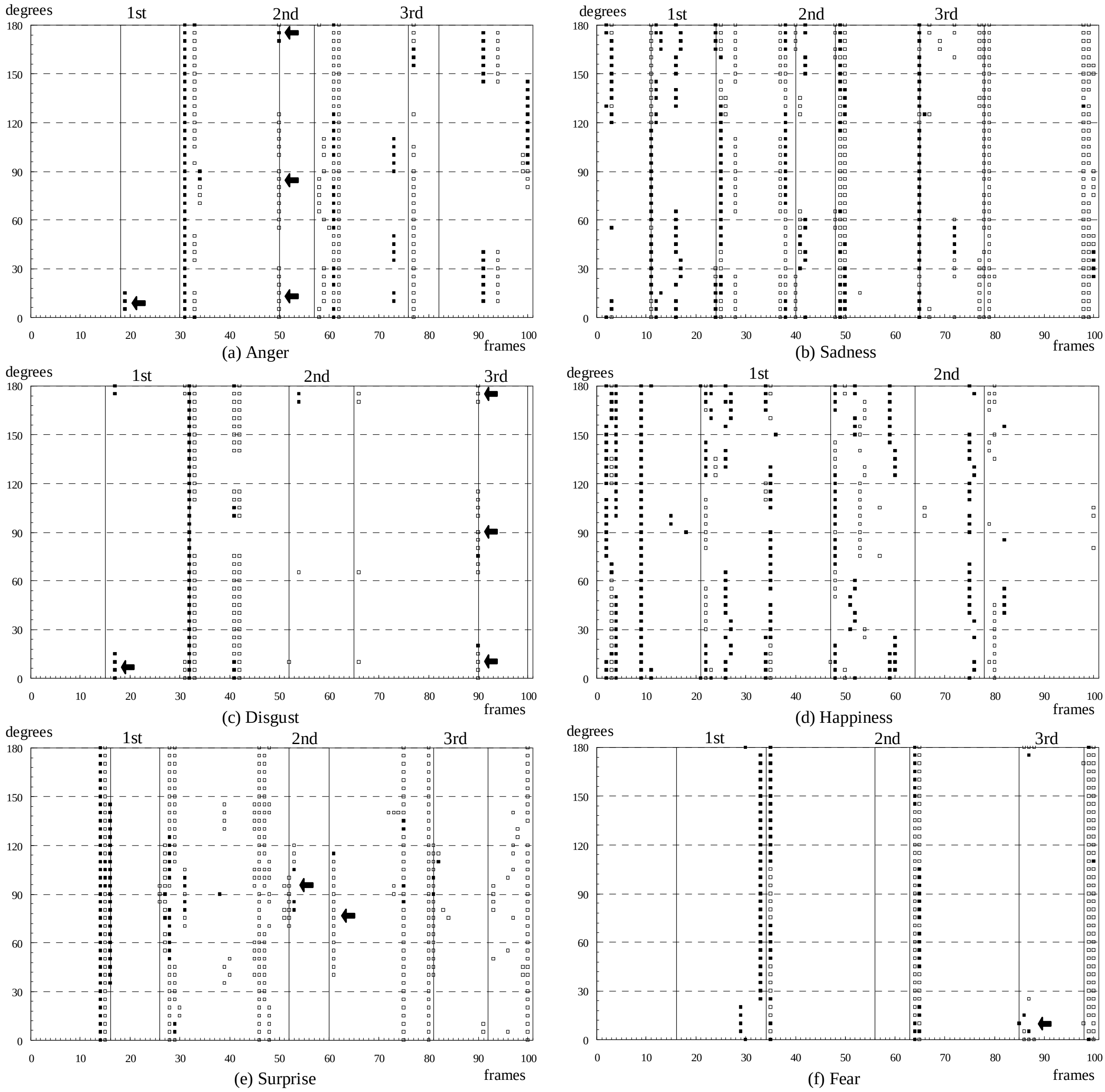


Figure 6. Categorical changes of ART2 networks for ρ = 0.970 in each facial expression: (a) $\sim$ (f). Filled rectangles represent the generation of new categories; empty rectangles represent transitions to existing categories. Vertical lines in each graph show the appearance or disappearance of facial expressions that correspond to Table I. The arrows show the point orientation selectivity represented.

although slight orientation selectivity is apparent. The categorical changes of happiness are redundant (Fig. 6d). The classification granularity seems to be short because the second expression can only detect two directions: 100 and 105 degrees. In this case, the setting value of ρ cannot be increased. The open level of the mouth differs before and after the 34th frame of the first appearance. That difference is detectable for the category changes. The open level of the mouth at surprise is characteristic (Fig. 6e). The categorical changes are noticeable around 90 degrees. However, in expectation of facial appearance points, the result is strongly reflective of eye blinking. The categorical changes of fear are not detected (Fig. 6f). This result indicates that the vigilance parameter, ρ = 0.970, is small.

Next, Fig. 7 shows the results of sadness, fear, and anger with changing vigilance parameters. The vigilance parameter of sadness were turned down step by 0.010 because the category is redundant in Fig. 6b. The vigilance parameter of fear and anger were turned up to 0.980 because the classification granularity is short in Figs. 6c and 6f. The redundant categories were decreased at the result of sadness in case of ρ = 0.960 (Fig. 7a). The category changes are seen at the first appearance. Moreover, in the case of ρ = 0.950, the category changes are more apparent at the first and second appearance (Fig.

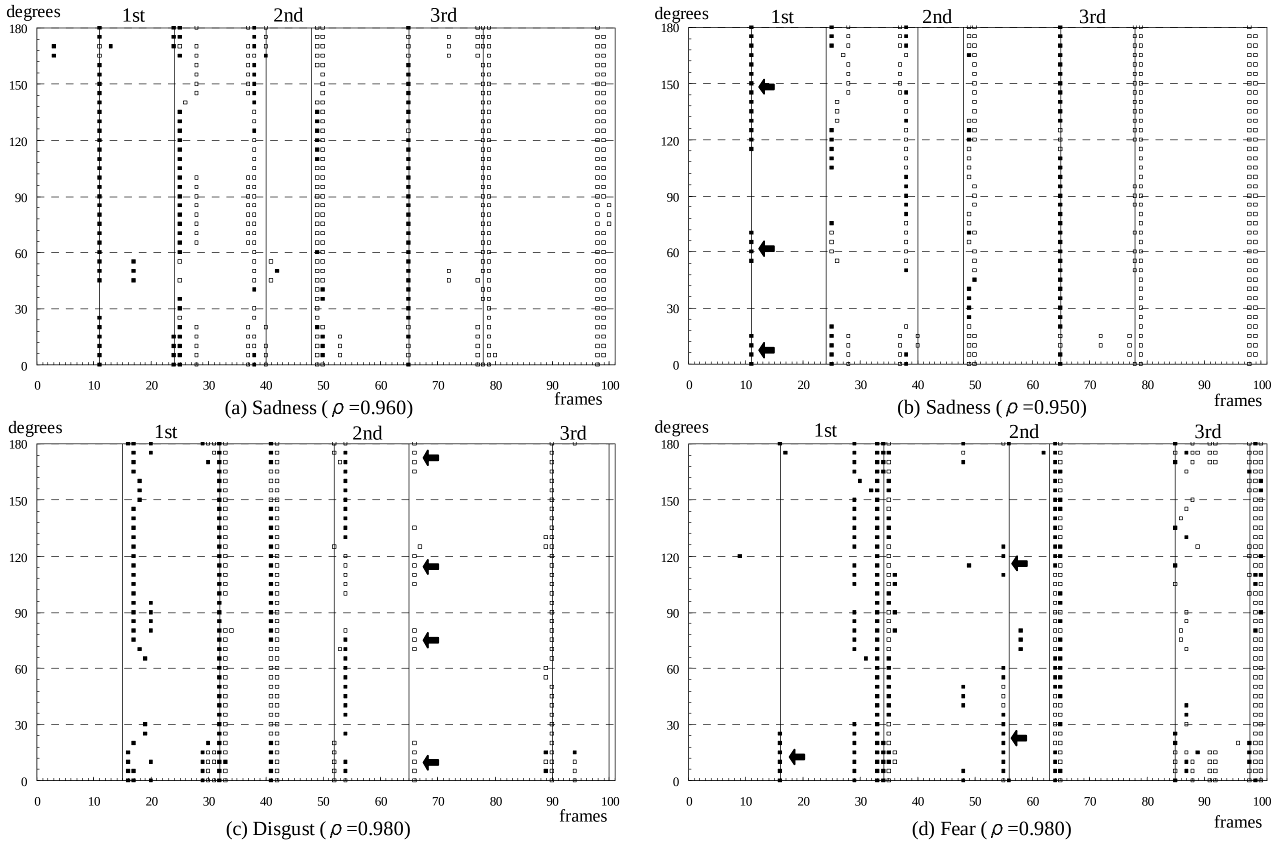


Figure 7. Categorical changes of ART2 networks ($\rho = 0.960, 0.950$ for sadness and $\rho = 0.980$ for disgust and anger). The arrows show the point orientation selectivity represented.

7b). The category changes appeared all directions at the 98th and 99th frames. The cause is eye blinking, which occurred in the frames. We consider that the features of eye blinking are easy to divide into other features because eye blinking occurred in almost all directions. The appearance and disappearance of disgust appeared in the case of ρ = 0.980, especially in the second one (Fig. 7c). In the case of ρ = 0.980 of fear, all appearances were detected (Fig. 7d). Especially, it was detected in a wide range at the second appearance.

The method can detect facial appearance points, even in the lower setting of ρ using orientation selectivity. In other words, the method can reduce redundant categories with the lower setting of ρ. Moreover, the method can detect facial expression changes with the range of avoiding redundant categories to increase the setting of ρ within orientation selectivity if the classification granularity is insufficient for a problem to be solved. We consider that the method can realize an advanced type of facial expression recognition for the next step of facial expression classification using the patterns of category changes with orientation selectivity.

V. Conclusion

This paper presents a method for representation of facial expression changes using orientation selectivity of Gabor wavelets on ART networks. The method produced suitable vigilance parameters according to classification granularity using orientation selectivity. Moreover, the method represented the appearance and disappearance of facial expression changes to detect dynamic, local, and topological feature changes from whole facial images.

Future studies must evaluate other response selectivity, such as wavelength, amplitude, frequency and direction of motion. In addition, we are going to take examinations about the formation of categories for long-term facial changes, implementation of oblivion mechanisms, fusion with context information, etc. to realize a natural and flexible MMI.

In our method, we selected the best size of category maps. The suitable training data are different in each problem to be solved. Automatic setting of the size of category maps is the subject of our future work. Moreover, we will apply our method to large-scale problems.

Acknowledgment

This work was supported by a Grant-in-Aid for Young Scientists (B) No. 21700257 from the Ministry of Education, Culture, Sports, Science and Technology of Japan.

References

[1] M. Pantic and L.J.M. Rothkrantz, “Automatic Analysis of Facial Expressions: The State of the Art,” *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 22, no. 12, pp. 1424–1445, Dec. 2000.

[2] M.J. Lyons, J. Budynek, and S. Akamatsu, “Automatic Classification of Single Facial Images,” *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 21, no. 12, pp. 1357–1362, Dec. 1999.

[3] M. Nishiyama, H. Kawashima, T. Hirayama, and T. Matsuyama, “Facial Expression Representation based on Timing Structures in Faces,” *IEEE Int’l. Workshop on Analysis and Modeling of Faces and Gestures*, pp. 140–154, 2005.

[4] P. Ekman and W. V. Friesen, “Unmasking the Face: A Guide to Recognizing Emotions from Facial Clues,” Malor Books, 2003.

[5] G.A. Carpenter and S. Grossberg, “ART 2: Stable Self-Organization of Pattern Recognition Codes for Analog Input Patterns,” *Applied Optics*, vol. 26, pp. 4919–4930, 1987.

[6] G.A Carpenter and S. Grossberg, “Pattern Recognition by Self-Organizing Neural Networks,” The MIT Press, 1991.

[7] G. Donato, M.S. Bartlett, J.C. Hager, P. Ekman, and T.J. Sejnowski, “Classifying Facial Actions,” *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 21, no. 10, pp. 974–989, Oct. 1999.

[8] D.H. Hubel and T.N. Wiesel, “Functional Architecture of Macaque Monkey Visual Cortex,” *Proc. Royal Soc. B (London)*, vol. 198, pp. 1-59, 1978.

[9] C. Liu, “Gabor-Based Kernel PCA with Fractional Power Polynomial Models for Face Recognition,” *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 21, no. 10, pp. 974–989, Oct. 1999.

[10] T.S. Lee, “Image representation using 2D Gabor wavelets,” *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 18, no. 10, pp. 959–971, Oct. 1996.

[11] A.K. Jain and S.K. Bhattacharjee, “Address block location on envelopes using Gabor filters: supervised method,” *11th IAPR International Conference on Pattern Recognition*, vol. II, pp. 264–267, Sep. 1992.

[12] C.P. Papageorgiou, M. Oren, and T. Poggio, “A general framework for object detection,” *Proc. International Conference on Computer Vision*, pp. 555–562, 1998.

[13] T. Randen and J.H. Husoy, “Filtering for texture classification: a comparative study,” *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 21, no. 4, pp. 291–310, Apr. 1999.

[14] D. Shan and R.K. Ward, “Statistical Non-Uniform Sampling of Gabor Wavelet Coefficients for Face Recongnition,” *Proc. IEEE International Conference on Acoustics, Speech, and Signal Processing*, vol.2, pp.73–76, 2005.

[15] C. Liu and H. Wechsler, “Gabor Feature Based Classification Using the Enhanced Fisher Linear Discriminant Model for Face Recognition,” *IEEE Trans. Image Processing*, vol.11, no.4, pp.467–476, 2002.

Hirokazu Madokoro received the ME degree in information engineering from Akita University in 2000 and joined Matsushita Systems Engineering Corporation. He moved to Akita Prefectural Industrial Technology Center and Akita Research Institute of Advanced Technology in 2002 and 2006, respectively. He received the PhD degree from Nara Institute of Science and Technology in 2010. He is currently an assistant professor at the Department of Machine Intelligence and Systems Engineering, Akita Prefectural University. His research interests include in machine learning and robot vision. He is a member of the Robotics Society of Japan, the Japan Society for Welfare Engineering, the Institute of Electronics, Information and Communication Engineers, and the IEEE.

Kazuhito Sato received the ME degree in electrical engineering from Akita University in 1975 and joined Hitachi Engineering Corporation. He moved to Akita Prefectural Industrial Technology Center and Akita Research Institute of Advanced Technology in 1979 and 2005, respectively. He received the PhD degree from Akita University in 1997. He is currently an associate professor at the Department of Machine Intelligence and Systems Engineering, Akita Prefectural University. He is engaged in the development of equipment for noninvasive inspection of electronic pats, various kinds of expert systems, and MRI brain image diagnostic algorithms. His current research interests include in biometrics, medical image processing, facial expression analysis, computer vision. He is a member of the Medical Information Society, the Medical Imaging Technology Society, the Japan Society for Welfare Engineering, the Institute of Electronics, Information and Communication Engineers, and the IEEE.