Facial Expression Spatial Charts for Representing of Dynamic Diversity of Facial Expressions

Hirokazu Madokoro, Kazuhito Sato, Akira Kawasumi, and Sakura Kadowaki
Faculty of Systems Science and Technology,
Akita Prefectural University
84–4 Aza Ebinokuchi Tsuchiya, Yurihonjo City, 015–0055 Japan
madokoro@akita-pu.ac.jp

***Abstract*—This paper presents a method to generate individual Facial Expression Spatial Charts (FESC) using Self-Organizing Maps (SOM) and Fuzzy Adaptive Resonance Theory (ART) networks. We specifically examine the dynamic diversity of facial expressions in time-series facial images after conversion using Gabor wavelet filters. The proposed method consists of three steps: the first step is to extract topological features from time-series facial image datasets using SOMs; the second step is to integrate weights of SOM into categories using Fuzzy ART networks; the third step is to create FESCs integrated by all arousal levels produced from categories of facial expressions in each basic facial expression. For considering the influence that stress gives an expression, we measured the psychological emphasis that a subject feels at that time. The result shows a negative correlation for psychological stress and the expanse of FESC, which means that the expression became poor during feelings of stress.**

***Index Terms*—Facial expression spatial charts, Arousal levels, SOM, Fuzzy ART, SRS-18.**

I. Introduction

For human communication, we use sound, words, speaking, and language to express verbal information. On the other hand, we can understand intentions and feelings from information of faces and facial expressions. Especially for people with whom we share a close relation, we can feel and understand health conditions or moods directly from facial expressions. Human faces send information of various types. We acquire nonverbal information visually to use for our rich communication [1].

In the 1970s, from a study of how to express emotions related to facial expressions, Ekman and Friesen defined six basic emotions (happiness, disgust, surprise, sadness, anger, and fear) and six facial expressions created by those six basic emotions that are universal among cultures [2]. They described that these are basic facial expressions because their associated emotions can be distinguished with high accuracy. Moreover, they proposed a Facial Action Coding Systems (FACS) as a method to describe facial expression changes from movements on a facial surface for use in behavioral sciences and psychology. The FACS was developed originally as a tool to measure facial expressions consisting of anatomical stand-alone Action Units (AUs). Globally, the FACS is the most popular and standard method to describe facial expressions objectively. It is useful to realize natural and flexible man-machine interfaces in the fields of human cognition and behavior science studies.

Real expressions are blended intermediate facial expressions that often show mixtures of two or three emotions. We often express various facial expressions simultaneously. For example, eyes can express crying but the mouth can express a smile when someone is moved by an extremely kind deed. The processes of expressive facial expressions differ among people just as the shapes of faces differ among people. For example, the range within which expressions on the facial surface change from an emotion differs among people. Regarding this difference, Akamatsu described facial diversity of two types [3]. Facial components such as eyes, eyebrows, and the mouth differ for each person. Facial features of those facial components' position, size, location, etc. also differ. This is called static diversity. However, we move facial muscles to express internal emotions unconsciously or express emotions as a message. Facial expressions are produced by facial components and their transition from a neutral facial expression. This is called dynamic diversity. Regarding facial recognition in the field of facial image processing, only the use of static diversity is sufficient to obtain good results. Facial expression recognition requires not only static diversity but also dynamic diversity as a time series to cope with facial pattern transitions.

This paper presents a novel method to describe subject-specific patterns from topological changes of facial expressions using Self-Organizing Maps (SOM) that contain self-mapping characteristics and Adaptive Resonance Theory (ART) that entails incremental learning for time-series data from whole facial images as appearance-based methods without complex feature points. Moreover, we discuss influences of a subject's psychological stress on facial expression changes.

II. Related Work

As a study to treat facial expression changes and its timing factors, Bassili [4] described the possibility of classifying facial expressions using movements of feature points that are captured by installed markers on a facial surface. However, the influence of components of movements is not clear because this method can not control stimulations of facial movements in a visual psychological experiment. In a recent study specifically examining dynamic diversity, Ohta et al. [5] proposed a method based on a model of facial structure elements. They pointed out that methods to detect overall movements of the whole

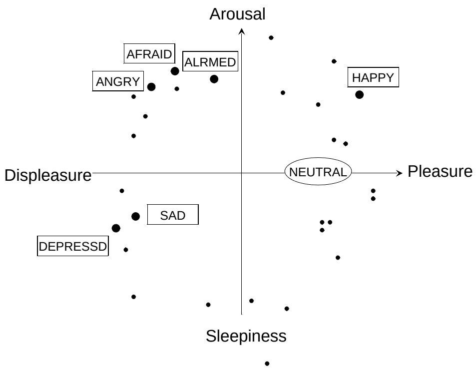


Fig. 1. Russell's circumplex model.

face can not detect fine and local variations of individual facial components such as eye blinking or changes in the shape of the mouth according to speech. Hirayama et al. [6] proposed facial scores for interpreting facial expressions based on temporary structures among partial movements in facial image sequences. They described many facial expressions that cannot be described by AUs because AUs are classified subjectively by an observer. In contrast, AUs are set as small units that are useful to classify facial expression movements visually and which stand alone anatomically. Although both methods are classifiable as feature-based methods, they use original setting feature points, not AUs. However, we humans can recognize facial expressions to infer local components and their movements from overall structures of a face when we infer intentions or emotions from a facial expression of a person. In fact, we extract not only movements of facial features such as the eyes, nose, mouth, and ears, but also local facial expression changes as topological changes from the whole face. Therefore, time-series facial movements play a role in understanding facial expressions. Moreover, we think the system can analyze facial expressions similarly to the human visual and cognitive system, which can examine a whole face that is changed overall as a unit of facial expression changes.

III. Facial expression spatial charts

Facial expression spaces are spatial configurations of each facial expression to regard semantic and polar characteristics of various emotions to be recognized by facial expressions. They represent a correspondence relation between physical parameters that present facial changes expressed by facial expressions and emotional parameters that are recognized as emotions. In this study, we define Facial Expression Spatial Charts (FESCs) as integrated facial expression spaces showing arousal levels in three facial expressions: happiness, anger, and sadness. The FESC is allocated on the axis of the pleasure level on Russell's circumplex model [8] shown in Fig. 1. We use FESCs to compose three facial expressions (happiness, anger, and sadness) in each arousal level created by each face shown in Fig. 2. The arousal level of facial expressions shows a quantized value of geometric topological changes according to facial expression changes from numerous facial expression patterns. For example, a bashful smile or a big smile being adopted to replace a neutral facial expression signifies happiness. The arousal level equals the number of categories after learning ART. The range of the arousal level is from the minimum level of 0–15 as a result of learning SOM and ART. An arousal level of zero represents a neutral facial expression of a subject. The arousal level is increased according to expansion of the richness and topological changes of facial expressions.

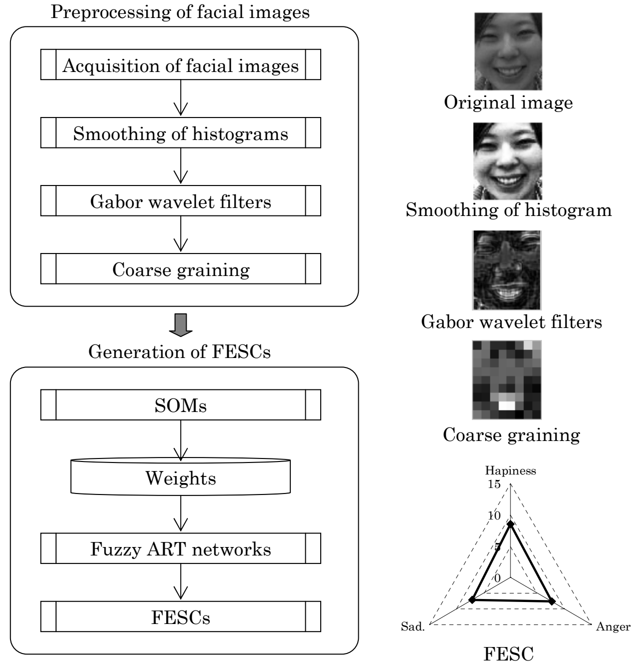


Fig. 2. Procedure of the proposed method from acquisition of facial images to generate FESCs.

IV. Proposed method

The goal of this study is representation of subject-specific facial expression spaces, particularly addressing the dynamic diversity of facial expressions. We propose an FESC as a chart to be described of subject-specific facial expression spaces using a SOM, which contains self-mapping characteristics and ART networks that entail incremental learning for time-series data. We address dynamic changes of facial expressions as topological changes of facial patterns driven by facial muscles of expression. We organize and visualize subject-specific facial expression spaces with arousal levels in each facial expression. Fig. 2 portrays procedures from the step of capturing time-series facial images to the step of creating FESCs. Detailed procedures describe the following.

A. Preprocessing of target images

The region of 80×90 pixels, including eyebrows, eyes, the nose, mouth, cheeks, and jaw, which all contribute to the impression of a whole face as facial feature parts, is extracted to a target for processing. After normalizing to grayscale images, structural features of local regions of visual patterns are extracted using Gabor wavelet filters that show similar characteristics to those of a human primary visual cortex.

Images are compressed to 72 dimensional spaces of 8×9 dimensional for coarse graining of facial features.

B. Classification of facial patterns with SOMs

We used SOMs with the fixed number of the Kohonen layer for normalizing facial expression patterns. The SOM proposed by T. Kohonen performs self-learning from features of input data [7]. The SOM maintains a topological information of input data as weights that are mapped to the Kohonen layer. Especially, the SOM has an excellent capability for visualizing input data as weights with vector compression. Therefore, SOMs are applied in various fields, especially in image processing.

The SOM is an unsupervised neural network with competitive learning in neighborhood regions. We use SOM with a one-dimensional mapping layer. The training algorithm of SOMs is the following.

1) Let $w_{i,j}(t)$ be the weight from the input unit i to the Kohonen unit (n,m) at time t. The weights are initialized with random numbers.
2) Let $x_i(t)$ be the input data to the input unit i at time t. The Euclidean distance d_j between $x_i(t)$ and $w_{i,j}(t)$ is calculated as
$$d_j = \sqrt{\sum_{i=1}^{I}(x_i(t) - w_{i,j}(t))^2}. \quad (1)$$
3) The win unit c is defined, for which d_j becomes a minimum as
$$c = argmin(d_j). \quad (2)$$
4) Let $N_c(t)$ be the units of the neighborhood of the unit c. The weight $w_{i,j}(t)$ inside $N_c(t)$ is updated using the Kohonen training algorithm as ($\alpha(t)$ is the training coefficient, which decreases with time.)
$$w_{i,j}(t+1) = w_{i,j}(t) + \alpha(t)(x_i(t) - w_{i,j}(t)). \quad (3)$$
5) Training is finished when the iterations reach the maximum number.

C. Integration of facial patterns with Fuzzy ART networks

We apply Fuzzy ART to integrate a suitable number of categories without pre-setting of the number of categories. Carpenter and Grossberg proposed ARTs of various types: ART1, ART1.5, ART2, ART2-A, ART3, ARTMAP, Fuzzy ART, Fuzzy ARTMAP, etc. [9]. We use Fuzzy ART [10], into which analog values can be input, which was proposed in 1991. The network consists of two fields: Field 1 (F1) for feature representation and Field 2 (F2) for category representation.

The Fuzzy ART algorithm is the following. I is an input m-dimensional vector. The numbers of neurons of the F1 and F2 are, respectively, M and N. Fuzzy ART dynamics are determined using a choice parameter $a(a > 0)$, a learning rate parameter $r(0 \leq r \leq 1)$, and a vigilance parameter $p(0 \leq p \leq 1)$.

1) w_i are the weights between each F2 neuron i and each corresponding F1 neuron. All w_i are initialized as one.
2) For each input I and each neuron i, the choice function T_i is defined as
$$T_i = \frac{|I \wedge w_i|}{a + |w_i|}, \quad (4)$$
where the fuzzy AND operator is defined as
$$(n \wedge v)_j \equiv min(u_j \wedge v_j), \quad (5)$$
and where the norm is defined as
$$|u| \equiv \sum_{j=1}^{m} |u_j|. \quad (6)$$
3) i_0, which is the maximum value of T_i, is selected for a category as a winner. The category with the smallest index is chosen if more than T_i is maximal. When i_0 is selected for a category, the i_0 th neuron on the F2 is set to 1 and other neurons are set to zero.
4) Resonance or resetting is judged as 5) if the selected category at 2) and 3) matches the input data I.
5) Resonance occurs if the match function of the chosen category meets the vigilance criterion. The weight vector w_{i0} is updated as
$$w_{i0} = r(I \wedge w_{i0}) + (1 - r)w_{i0}. \quad (7)$$
6) If I does not have resonance to i_0, then i_0 is reset. The network seeks a next category T_i to be maximal and reselects it. The network determined resonance or reset. If all categories are reset, then go to 7).
7) A neuron is created on F2 and a new category is registered. Steps 2)–7) are controlled using M and K and are repeated $M \times K$ times to be presented sequentially of I.

V. Experimental results

A. Facial expression datasets

Open datasets of facial expression images are generally employed in many conventional studies. In static facial images, the dataset presented by Ekman and Friesen is the most popular dataset comprising collected various facial expressions used for visual stimulation in psychological examinations of facial expression cognition. As dynamic facial images, the Cohn–Kanade dataset and the Ekman–Hager dataset are widely used, especially in experimental applications. However, numerous facial images are necessary for evaluation and comparative experiments using both static and dynamic images. Moreover, the sample images that are taken for the same person and the same facial expression under controlled conditions are too few, presenting limitations for their experimental use. Certainly, original datasets that are collected by researchers and each organization are used in many studies.

We took facial images at one-week intervals for 10 subjects, including five female university students (Subjects A, B, C, and D were 19; and Subject E was 21) and five male university students (Subjects F and J were 19; and Subjects G, H, and I were 22). The purpose of this capture of horizontal datasets from 10 subjects and perpendicular datasets for each subject

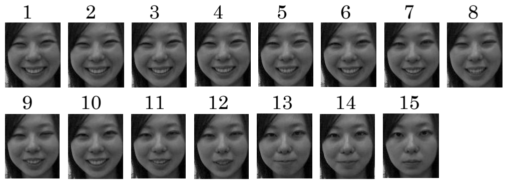


Fig. 3. Weighted images as SOM training results.

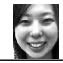
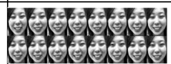
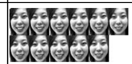
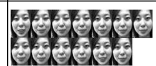
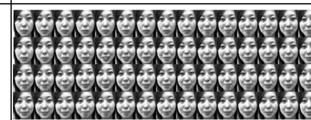
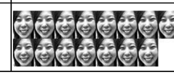
Category	Classification results by SOM (Weighted Images)	Integrated units	Ave.
1		2,3	
2		5,6	
3		7,8,9	
4		10,11	
5		12	
6		13,14,15	
7		1	
8		4	

Fig. 4. Weighted images of each unit and integration results with Fuzzy ART networks.

is to address individual differences between subjects and time-dependent changes in each subject. We set the environment to simulate normal indoor conditions (lighting by fluorescent lamps). We took frontal facial images to include the region containing facial components such as the eyebrows, eyes, the nose, and the mouth. Our target facial expressions are happiness, anger, and sadness that a subject expresses intentionally. We took each facial expression of a subject with 8-bit gray-scale images of 200 frames (20 s by 10 frames per second).

B. Stress measurements

We use Stress Response Scale 18 (SRS-18) [12] sheets for measuring the stress of a subject. The SRS-18 can measure responses related to psychological stress easily in a short time and record many that we meet in our daily life. Specific psychological stress responses are gloom, anxiety, and anger (emotional responses), lethargy and difficulty concentrating (cognitive responses), decreased efficiency of work (behavioral responses), etc. caused by stressors. Stressors mean stimuli that are caused by stress. This sheet can measure stress responses according to three factors: dysphoria or anxiety, displeasure or anger, and lassitude. There are 18 questions that can elicit answers of four types: strongly no, no, yes, and strongly yes. The scores for answers correspond respectively to zero to three points. The range of total points is 0–54 points. The total score is high, meaning that the psychological stress is high. We measure this stress before taking facial expression images. We showed no result of this check to a subject.

C. Classification results with SOMs

Each SOM is trained with the set of 200 static images extracted from time-series images of three facial expressions (happiness, anger, and sadness). We set the Kohonen layer to 15 units for learning dynamic and topological changes of facial expressions. The advantage of using SOM is to learn topological characteristics of high-dimensional facial images to compress to low-dimensional vectors. The SOM can learn characteristics of topological information of facial expressions for clustering to 15 units as a mapping space from 200 facial expression images. We obtained mapping results of SOM to change the Kohonen layer from 9 units to 21 units in steps of 2 units. For large quantities of units, unburst units appear because the mapping space is too large to represent topological changes of facial expressions. On the other hand, in the case of fewer units, the topological information of faces is not reflected to the units. We set the Kohonen layer to 15 units as the minimum number of units according to these preexperimental results. In the next step, units that have closed weights are integrated by Fuzzy ART networks. We set the input layer to 72 units, which is the same size of the resolution of input images after coarse graining at the final step of preprocessing. We set the number of learning iterations of the SOM to 500. We confirmed that the SOM mapped topological information of facial images into weights as a result of learning of this number of iterations. In this step, the 200 images are classified into 15 categories, which is the same number of units on the Kohonen layer.

Fig. 3 presents the classification result of Subject A using the dataset during 20 weeks of happiness. The correlation values of weights of neighborhood units are similar. Therefore, similar facial images are close to each other to be shown in Fig. 3 because the SOM learns similar features of weights between units.

D. Integration results with Fuzzy ART

We used the weights of 15 units of the SOM as training data to the Fuzzy ART network. We set the F1 of the Fuzzy ART to 72 units, which is the same number of input units of the SOM. Fuzzy ART learns weights of SOM corresponding to each unit. Units with similar weights that are classified with SOM of 15 units are integrated as categories with Fuzzy ART.

The parameters of the Fuzzy ART respond very sensitively, especially the vigilance parameter, which controls the classification granularities. We set the vigilance parameter to 0.90. When we set the vigilance parameter to a low value, the granularity is low and the number of categories is decreased. However, when we set the vigilance parameter to a high value, the granularity is high and the number of categories is increased. For setting of a high value close to 1.0, the Fuzzy ART responds in small input values and produces many new categories. When the granularity is too high, Fuzzy ART

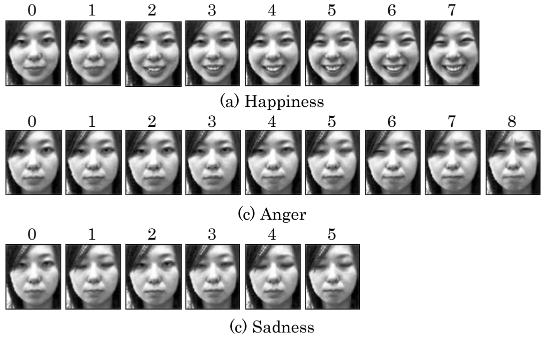


Fig. 5. Face images of each arousal level.

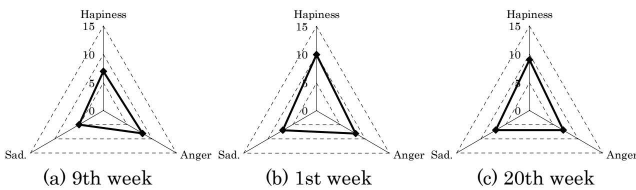


Fig. 6. Typical results of FESCs.

creates a new category for the input data that should categorize same category. However, our method conducts pattern classifications that are learned with similar features once in each unit after learning of SOM by self-mapping characteristics from topological changes of facial patterns. Therefore, we set the vigilance parameter to 0.90 from a preliminary experiment for integrating units with saving topological information after learning topological characteristics of facial expressions. Fig. 4 shows weighted images of each unit and integration results with Fuzzy ART networks. The left row in this figure shows average facial images in each category. Suitable quantities of categories were obtained for uncertain target problems using Fuzzy ART networks.

E. Generation results of FESC (Subject A)

We visualized facial expression spaces after sorting patterns of arousal levels from the category of neutral facial expression to each category of the average facial expression images. Fig. 5(a) shows facial expression images of happiness in each arousal level of Subject A. These images show subject-specific facial expression spaces of happiness after learning SOM and ART from the original 200 images. Our method quantizes arousal levels to eight steps from the neutral facial expression level zero to the maximum expression level seven. We extracted arousal levels of anger and sadness shown in Figs. 5(b) and 5(c) respectively as in the processing procedures to extract the arousal level of happiness. The arousal levels of anger and sadness are, respectively, 8 and 5. Fig. 6 portrays an FESC of a facial expression space to be integrated. The center of the FESC means that the arousal level is zero as a neutral facial expression. The arousal level is increased according to the distance outside of the triangle. Fig. 6(a) shows a FESC of the ninth week of a subject. Figs. 6(b) and 6(c) respectively show an FESC when the stress levels were the maximum and the minimum. We analyze these results in conjunction with

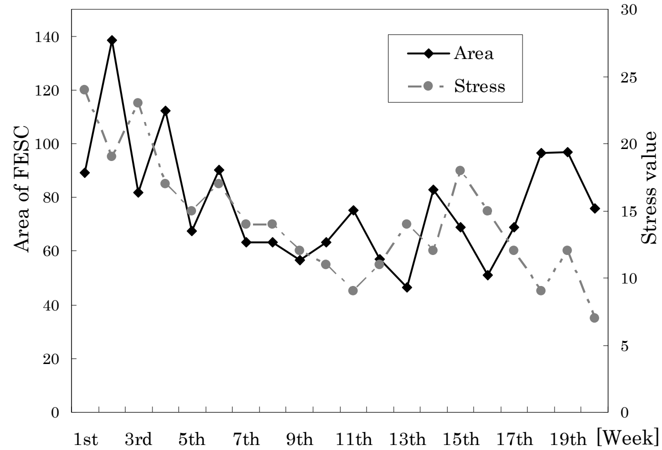


Fig. 7. Transition of FESC and stress during 20 weeks of Subject A.

stress analysis results obtained using SRS-18.

F. Relation between FESC and stress (Subject A)

Fig. 7 shows the areas of the FESC and the stress valued measured by SRS-18 for 20 weeks. The left side and the right side vertical axes respectively portray the area of FESCs shown in Fig. 6 and stress values according to the stress sheets. This area shows the quantification of FESCs of three facial expressions (happiness, anger, and sadness) that a subject expressed subjectively. The stress score is expressed as four steps from 1–4 points in three factors: dysphoria and anxiety, displeasure and anger, and lassitude. The total score is 54 points.

The areas of the FESC show high scores in the early weeks. According to the traces, these scores reach steady levels. One factor explaining this result is that the subject did not get used to having facial expressions taken in the early weeks. Fig. 6(b) show the FESC is decreasing when Subject A felt a stress strongly. Fig. 6(c) shows that the FESC is increasing when Subject A did not feel stress. Especially, we confirmed tendencies that appeared from the first week through the fourth week and from the ninth week through the eleventh week shown in Fig. 7. Generally, stress affects not only the body and acts; it also strongly affects facial expressions. A person who has stress has poor facial expression. The effects by stress on facial expressions vary among individuals because the perception and effects of stress differ among people. Therefore, we analyzed facial expression datasets over a long period. We also evaluated the effects between psychological stress and facial expressions, and time-dependent changes of facial expressions.

G. Results of 10 subjects

We applied our method to 10 subjects (Subjects A–J) to take facial expression datasets from 7 weeks to 20 weeks in each subject for tracing each FESC and stress. Fig. 8 presents correlation coefficients between stress values and the total area of FESCs and each area of each facial expression in each subject. The vertical axis shows correlation coefficients; the

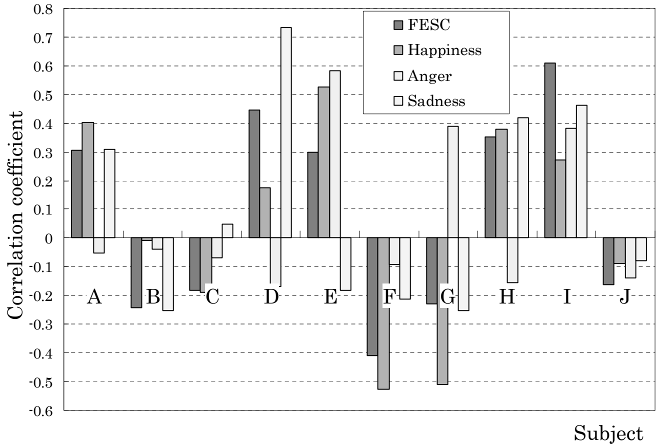


Fig. 8. Correlation coefficient between stress and each facial expression of 10 subjects.

TABLE I
FACIAL EXPRESSIONS AFFECTED BY STRESS IN 10 SUBJECTS

Person	Facial expression	Person	Facial expression
A	Happiness	F	Happiness
B	Sadness	G	Happiness
C	None	H	Sadness
D	Sadness	I	Sadness
E	Anger	J	None

horizontal axis shows the subjects. The positive direction of the vertical axis shows that the positive correlation coefficient means that the facial expression space is expanded and that the arousal level is higher because the stress is higher. The negative direction of the vertical axis shows that the negative correlation coefficient means that the facial expression space becomes narrow and that the arousal level is lower because the stress is higher. The near zero value of the vertical axis means that facial expressions are unaffected by stress.

Subject I is a typical case showing a positive correlation. All facial expressions of Subject I show positive correlations. The correlation coefficient between the FESC and stress is 0.61. Subject F is a typical case showing a negative correlation. Particularly, Subject F has a strong negative correlation in the facial expression aspect of anger. However, subjects C and J show no significant correlation coefficients in any facial expression. Particularly, Subject C can produce facial expressions with a steady number of facial expression patterns, independent of stress, although she exhibits strong stress with a high value.

Table I portrays the facial expression of the highest absolute value of the correlation coefficient as a facial expression that is strongly affected by psychological stress. Although facial expressions affected by stress differ among subjects, as an overall tendency, happiness and sadness are affected by stress but anger is unaffected by stress. The effect of stress on facial expressions differs among subjects because the perception of stress, its degree, and its effects on the body differ among subjects as a result of personality and body condition difference among individuals. Therefore, we must continue to analyze influences of stress effects on facial expressions of a subject and time-series changes of facial expressions over long periods.

VI. CONCLUSION

This paper presents a method to create FESCs using SOM and ART with emphasis on dynamic diversity of facial expressions from time-series facial images. Using brightness values of images after conversion using Gabor wavelet filters from dynamic topological changes produced by facial expression muscles, our method can represent subject-specific facial expression spaces using FESCs while maintaining topological information with appearance-based feature representations to represent features of the whole face without setting of feature points.

Future studies must evaluate natural facial expressions and intentional ones for discrimination using symmetry properties of the horizontal direction to represent facial expression rhythms created by individual patterns of time changes of arousal levels to increase the number of subjects, and to analyze long-term data for elucidating the relation between facial expressions and stress. Moreover, we will increase the number of subjects for horizontal studies.

ACKNOWLEDGMENT

The authors thank our fellow engineers at SmartDesign LLP for giving us much appreciated advice and for developing a system for capturing facial images for this experiment. We also thank the 10 students of our university who participated as subjects by letting us take facial images over a long period.

REFERENCES

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

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

[3] S. Akamatsu, "Recognition of Facial Expressions by Human and Computer [I] : Facial Expressions in Communications and Their Automatic Analysis by Computer," *The Journal of the Institute of Electronics, Information, and Communication Engineers*, vol. 85, no. 9, Sep. 2002, pp. 680-685.

[4] J.N. Bassili, "Facial motion in the perception of faces and of emotional expression," *Journal of Experimental Psychology: Human Perception and Performance*, vol. 4, no. 3, 1978, pp. 373-379.

[5] H. Ohta, H. Saji, and H. Nakatani, "Recognition of Facial Expressions Using Muscle-Based Feature Models," *The Transactions of the Institute of Electronics, Information and Communication Engineers*, D-II, vol. J82-DII, no. 7, Jul. 1999, pp. 1129-1139.

[6] T. Hirayama, H. Kawashima, M. Nishiyama, and T. Matsuyama, "Facial Expression Representation based on Timing Structures in Faces," *The Transaction of Human Interface Society*, vol. 9, no. 2, May 2007, pp. 271-281.

[7] T. Kohonen, Self-organizing maps, Springer Series in Information Sciences, 1995

[8] J.A. Russell and M. Bullock, "Multidimensional Scaling of Emotional Facial Expressions: Similarity From Preschoolers to Adults," *Journal of Personality and Social Psychology*, vol. 48, 1985, pp. 1290-1298.

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

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

[11] H. Madokoro, K. Sato, and M. Ishii, "Orientation Selectivity for Representation of Facial Expression Changes," *The International Joint Conference on Neural Networks*, vol. 12-17, Aug. 2007, pp. 1210-1215.

[12] S. Suzuki, "Stress Response Scale-18", Jul. 2007.