Estimation of Psychological Stress Levels Using Facial Expression Spatial Charts

Hirokazu Madokoro and Kazuhito Sato
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 new framework to describe individual facial expression spaces, particularly addressing the dynamic diversity of facial expressions that appear as an exclamation or emotion, to create a unique space for each person. We name this framework Facial Expression Spatial Charts (FESCs). The FESCs are created using Self–Organizing Maps (SOMs) and Fuzzy Adaptive resonance Theory (ART) of unsupervised neural networks. In the experiment, we created an original facial expression dataset consisting of three facial expressions—happiness, anger, and sadness—obtained from 10 subjects during 7–20 weeks at one-week intervals. Results of creating FESCs in each subject show that the method can adequately display the dynamic diversity of facial expressions between subjects, in addition to temporal changes in each subject. Moreover, we used stress measurement sheets to obtain temporal changes of stress in each subject for analyzing psychological effects of the stress that subjects feel. We analyzed relations between numerous individual facial expression patterns and psychological stress values. Results show that facial expressions when influenced by stress differ among subjects. Moreover, we estimated stress levels of four grades using Support Vector Machines (SVMs). The mean estimation rates for all 10 subjects and for 5 subjects over more than 10 weeks were, respectively, 68.6 and 77.4%.**

***Index Terms*—Facial Expression Spatial Charts, SOMs, Fuzzy ART, SVMs, SRS-18.**

I. Introduction

A face sends information of various types. Humans can recognize intentions and emotions from diverse information that is exhibited through 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. For the role of facial expressions in human communication, it is desirable to develop advanced interfaces between humans and machines in the future [1].

In the 1970s, from a study to determine how to express emotions related to facial expressions, Ekman and Friesen defined six facial expressions shown by people feeling six basic emotions (happiness, disgust, surprise, sadness, anger, and fear) that are apparently universal among cultures. They described that these are basic facial expressions because their associated emotions are distinguishable with high accuracy. However, real expressions are blended intermediate facial expressions that often show mixtures of two or three emotions. Human beings 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. Moreover, the processes of expressive facial expressions contain individual differences such as differences of face shapes among people.

Regarding this difference, Akamatsu described that human faces present diversity of two types: static diversity and dynamic diversity [3]. Static diversity is individual diversity that is configured by facial componential position, size, location, etc., consisting of the eyes, nose, mouth, and ears. We can identify a person and determine their gender and impressions using static diversity. We are able to move facial muscles to express internal emotions unconsciously and sequentially or express emotions as a message. This is called dynamic diversity. Facial expressions are expressed as a shift from a neutral facial expression to one of changed shapes of parts and overall configurations constructed with the face. For studying facial expression analysis, we must consider and understand not only static diversity but also dynamic diversity.

For organizing and visualizing facial expression spaces, this paper presents a novel framework to describe the dynamic diversity of facial expressions. The framework accommodates dynamic changes of facial expressions as topological changes of facial patterns driven by facial muscles of expression [4]. For that reason, the framework is suitable to describe the richness of facial expressions using Arousal Levels (ALs). The target facial expressions are happiness, anger, and sadness from the basic six facial expressions to represent expression levels as a chart with axes of each expression quantitatively and visually. From temporal facial expression images, we use Self-Organizing Maps (SOMs) that contain self-mapping characteristics to extract facial expression categories according to expressions. We also use Adaptive Resonance Theory (ART) that contains stability and plasticity that enable classification to integrate categories adaptively under constant granularity. We infer relations between categories created by Fuzzy ART and ALs based on Russell's circumplex model. This paper presents Facial Expression Spatial Charts (FESCs) to represent dynamic diversity of facial expressions as a dynamic and spatial chart. For the experiment, we created original facial expression datasets including images obtained during 7–20 weeks at one-week intervals from 10 subjects with three facial expressions: happiness, anger, and sadness. Experimental results show that our method can visualize and quantify facial expressions between subjects and temporal changes for

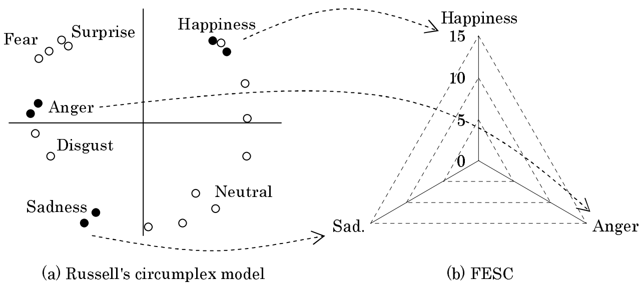


Fig. 1. Correspondence relations between Russell's circumplex model and FESC.

creating FESCs in each subject. We use stress measurement sheets to assess temporal changes of stress in each subject for analyzing psychological stress, which includes the subjects that affect facial expressions. We analyze relations between FESCs and psychological stress values. Moreover, we estimate stress levels from FESCs.

II. Arousal Levels and Facial Expression Spatial Charts

As described in this paper, we introduce ALs as an index of quantification of facial expression spaces. ALs show quantized values of the arousal dimension of the vertical axis on the Russell's circumplex model [5] portrayed in Fig. 1(a). All emotions are constellated on a two-dimensional space: the pleasure dimension of pleasure–displeasure and the arousal dimension of arousal–sleepiness. As described in this paper, we specifically examine the arousal dimension on Russell's circumplex model. We define ALs as a quantized value of topological and geometrical changes of facial patterns from the neutral facial expression, which is the basis of each facial expression space. We consider that we are able to extract the dimension of ALs from facial expression images because we are examining intentional facial expressions. When creating facial expressions, humans move facial muscles irrespective of pleasure or displeasure while maintaining a certain mental status, although elements of the pleasure dimension are included in the expression. Therefore, as experiments or datasets to examine intentional facial expressions, the expression patterns are strongly correspondent to the arousal dimension. In contrast, the pleasure dimension is evaluated using a stress sheet that is often used in the field of psychology because it is difficult to address it in intentional facial expressions used for our experiment.

Facial expression spaces are spatial configurations of each facial expression that are used to analyze semantic and polar characteristics of various emotions portrayed by facial expressions [3]. They represent a correspondence relation between the physical parameters that present facial changes expressed by facial expressions and the psychological parameters that are recognized as emotions. Psychological parameters can be extracted from psychological experiments to take cognitive decisions related to emotions. Physical parameters must be described based on a certain standard of types and based on the

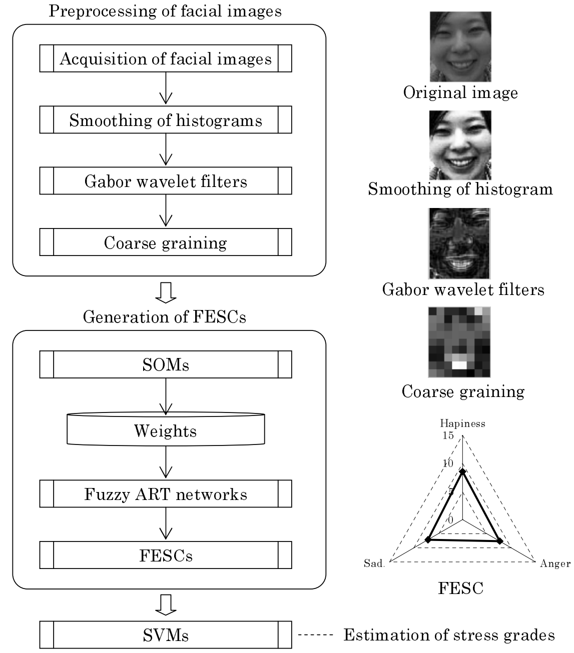


Fig. 2. Procedure of the proposed method from acquisition of facial images to generate FESCs.

facial deformity that invariably arises from expressions on different facial patterns that differ in each person, as represented by FACS. This paper presents Facial Expression Spatial Charts (FESCs) as a new framework to describe facial expression spaces and patterns of ALs constituting each facial expression. As described in this paper, our target facial expressions are happiness in the first quadrant, anger in the second quadrant, and sadness in the third quadrant of Russell's circumplex model. Fig. 1 shows the correspondence relation between Russell's circumplex model and an FESC. The value of each axis on the FESC shows the maximum values of ALs. The FESC is created by the connection among maximum values of ALs.

III. Proposed method

Akamatsu described the adaptive learning mechanisms necessary for modification according to individual characteristic features of facial expressions because the processes of expression differ among individuals. For example, a subject expresses facial surface changes of a certain size; the expressions and their sizes differs among individuals because the shapes of faces differ among people. Therefore, in this study, our target is intentional facial expressions of a person. We use SOMs for extracting topological changes of expressions and normalizing that are compressed in the direction of the temporal axis. In fact, SOMs perform unsupervised classification input data into a mapping space that is defined preliminarily. After classification by SOMs, facial images are integrated using Fuzzy ART, which is an adaptive learning algorithm with stability and plasticity. Fuzzy ART performs unsupervised classification at a constant granularity that is controlled by the vigilance parameter. Therefore, using SOMs and Fuzzy ART, time-series datasets showing changes over a long term are classified with a certain standard. Fig. 2 depicts an overview of the procedures used for our proposed method. Detailed procedures of Feature extraction, category classification with SOMs, category integration with Fuzzy ART, and stress estimation with SVMs are explained below.

A. Feature extraction

For this study, we use view-based feature representation of holistic images, not feature-based representation such as AUs. Actually, feature-based representation is superior to view-based representation for detailed description of local feature changes related to expressions. In contrast, feature-based representation demands high calculation costs for the process of extracting and tracing feature points. Moreover, feature-based representation contains problems of precision and stability in cases of numerous samples being processed automatically. For our method, we use view-based representation after converting images with filters of Gabor wavelets showing similar characteristics to those of a human primary visual cortex. Our processing target is to extract ALs from pattern changes of one facial expression from neutral facial expression. Therefore, we consider that the changed parts are apparent on the feature space after converting Gabor wavelets, without tracking of feature points based on AUs

The period during which images were obtained was expanded from several weeks to several months. We were unable to constrain external factors completely, e.g. through lighting variations, although we took facial expression images in constant conditions. Therefore, in the first step, brightness values are preprocessed with normalization of the histogram to the target images. In the next step, features are extracted using Gabor wavelet filtering. In the field of computer vision and image processing, information representation of Gabor wavelets is a popular method for an information-processing model based on human visual characteristics. The information representation of Gabor wavelets that can emphasize an arbitrary feature with inner parameters shows the same characteristic of response selectivity in a receptive field.

At the final step, we applied downsampling for noise reduction and compression of the data size. In this method, we set the initial position of the template to contain facial parts for capturing facial images. We use template-matching methods to trace the region of interest of a face in real time However, the trace results of the region of interest yield errors caused by body motion. These errors can be removed through the procedure of downsampling. The downsampling window that we set is 10×10 pixels. The dimension of the target images is compressed from 80×90 pixels to 8×9 pixels.

B. Classification of facial patterns with SOMs

For classification according to ALs, 200-frame images are normalized in a constant range. In this method, we used SOMs, which are unsupervised neural networks with competitive learning in neighborhood regions. Fig. 3(a) depicts a network architecture of a SOM. The network architecture of SOMs typically includes two layers: the input layer and the mapping layer. All units on the mapping layer are connected to all units of the input layer while maintaining weights between both layers. When a set of input data is propagated, a unit whose weights are the most similar to the input data is burst. Weights on the burst unit and its neighbor units are updated to be close to the input data, which is the learning of SOMs. Similarity among input data limits the features of topological saving that are reflected in the distance of the burst unit on the one-dimensional or two-dimensional units. According to the progress of learning, similar feature weights are mapped to neighbor units; other units are mapped to separate units.

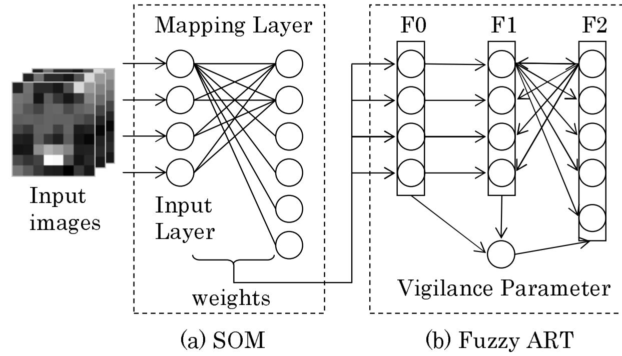


Fig. 3. SOMs and Fuzzy ART.

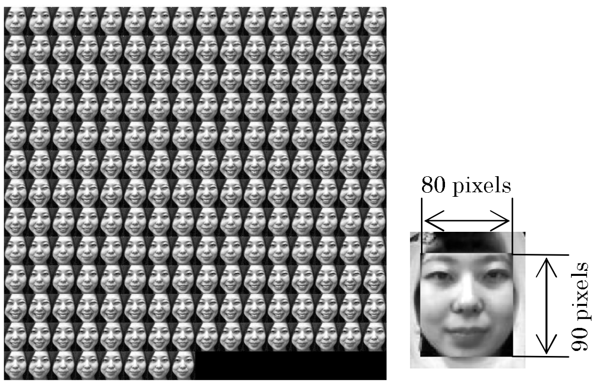
C. Integration of facial patterns with Fuzzy ART

The input data are classified in the fixed number of units of the mapping layer. Therefore, classification results are relative. In contrast, classification under the fixed granularity is required for long-term datasets in each subject. In our method, facial expression categories are integrated with Fuzzy ART to learn weights of SOMs.

The use of ART, which was proposed by Grossberg et al., is a theoretical model of unsupervised and self-organizing neural networks forming a category adaptively in real time while maintaining stability and plasticity. Actually, ART has many variations: ART1, ART1.5, ART2, ART2-A, ART3, ARTMAP, Fuzzy ART, Fuzzy ARTMAP, etc. [**?**]. We use Fuzzy ART [**?**], into which analog values can be input. Fig. 3(b) depicts a network architecture of SOMs. The network architecture of Fuzzy ART consists of three fields: Field 0 (F0) of receiving input data, Field 1 (F1) for feature representation, and Field 2 (F2) for category representation.

D. Allocation of ALs to FESCs

The facial expression categories classified by SOMs and integrated by Fuzzy ART are sorted in the order of ALs from the neutral facial expression category. For this dataset, the number of images of neutral facial expression is the maximum. The neutral facial expression category is selected to the maximum number of images. The ALs are sorted by correlation values in each category. The center of an FESC is AL 0, which represents a neutral facial expression. With increasing ALs, facial expression categories are assigned to the outside of the triangle.



(a) Acquired images (200 flames) (b) ROI

Fig. 4. Set of target images and regions of interest.

E. Facial expression dataset

Human show facial expressions of two types: spontaneous facial expressions and intentional facial expressions. Taking a steady and long-term dataset without regard to a camera and a situation is a challenging task, although spontaneous facial expressions present the advantage of corresponding directly to affection or emotion. Moreover, the cause-and-effect relation of facial expressions from emotions is uncertain. In contrast, intentional facial expressions are used as a communication method to communicate something positively to other person, especially in social communication. We set a target to create an original international facial expression datasets to obtain a long-term facial expression dataset for selected subjects. Moreover, intentional facial expression datasets are suitable to keep the number of subjects as a horizontal dataset.

Open datasets of facial expression images are released from some universities and research institutes to be used generally in many conventional studies for performance comparisons of facial expression recognition or automatic analysis of facial expressions [1]. However, the specifications vary in each dataset, and among datasets. As static facial images, the dataset presented by Ekman and Friesen is a 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 [7]. In recent years, the MMI Facial Expression Database presented by Pantic et al. [9] has become a widely used open dataset containing both static and dynamic images. These dynamic datasets contain a sufficient number of subjects as a horizontal dataset. However, images are taken only once for each person. No dataset exists in which the same subject has been traced over a long term. As described in this paper, we created an original dataset to take facial expression images over a long period in each subject as a vertical dataset.

F. Acquisition of facial expression images

We took images of three facial expressions with 10 subjects over a long term. The terms of taking images differed among subjects, but images were taken during 7–20 weeks at one-week intervals. Details of subjects are five females (Subjects A, B, C, and D were 19; Subject E was 21) and five males (Subjects F and J were 19; Subjects G, H, and I were 22) university students.

We began to take facial expression images when a subject became accustomed to the experimental environment after some trials. Considering generality and usability, we used a USB camera (Qcam, Logicool; Logitec Corp.). We set the environment to simulate a normal indoor condition (lighting by fluorescent lamps). We took frontal facial images to include the region containing facial components such as the eyebrows, eyes, nose, and mouth. We previously indicated to subjects to restrain the head position as much as possible. The images were fit to the constant range including the facial region. We used a method using Haar-like features and Boosting for tracking a face region to adjust the centers of images [**?**].

Fig. 4 depicts captured images. We set the region of interest to 80×90 pixels including the eyebrows, eyes, nose, mouth, cheeks, and jaw, which all contribute to the impression of a whole face as facial feature components. Our target facial expressions are happiness, anger, and sadness, which were all expressed intentionally. One set of datasets consisted of facial expression image sequences with neutral facial expression and each facial expression to be indicated. As an assumption, we instructed subjects to express an emotion 3–4 times during the image-taking time of 20 s. One set of data consisted of 200 frames with the sampling rate of 10 frames per second.

G. Stress measurements

For this study, we used stress measurement sheets known as the Stress Response Scale 18 (SRS-18) by Suzuki et al. [10]. The SRS-18 comprises question sheets that 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. This sheet can measure stress responses according to three factors: dysphoria or anxiety, displeasure or anger, and lassitude. The SRS-18 has 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. A high total score indicates a high level of psychological stress. Moreover, four grades of Level 1 (weak), Level 2 (normal), Level 3 (slightly high), and Level 4 (high) are classified from the points. Subjects complete this sheet before taking facial expression images to reduce the effect from stress checking results.

IV. Experimental results

A. Results of FESCs

Fig. 5 presents results of temporal changes of ALs of happiness, anger, and sadness. The horizontal axis depicts the frames that consist of 200 frames in each dataset. The vertical

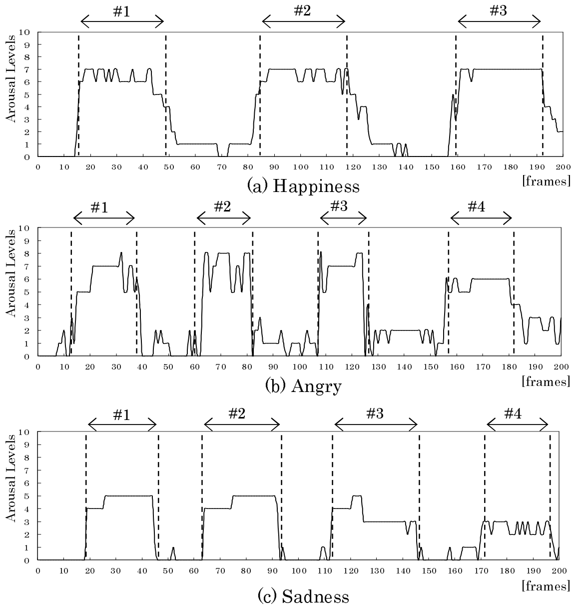


Fig. 5. Time-series changes of arousal levels (Subject A).

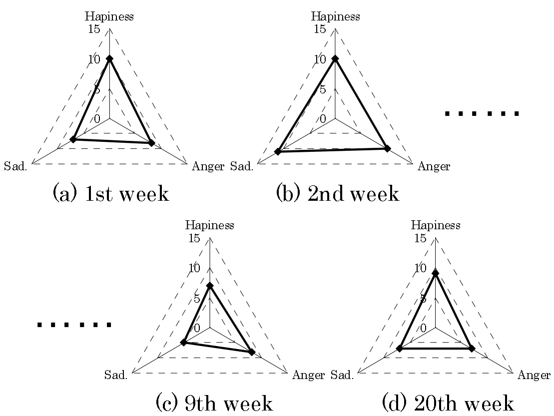


Fig. 6. Facial expression spatial Charts (Subject A).

axis depicts ALs. We marked the dashed vertical lines to the start and terminal positions of expression. The subject showed expressions of three or four times during one dataset. In this dataset of Subject A at the ninth week, happiness is expressed three times; anger and sadness are expressed four times. Start and terminal timings of expression are represented as changes of ALs. The ALs are changed according to the expressions, although the result contains slight variation. Fig. 6 depicts some examples of FESCs. The FESCs show temporal changes of facial expression patterns that changed in each week in the same subject. We consider that the changes are attributable to psychological effects. In the next, we will analyze these results with stress that is assessed as measured using the SRS-18.

B. FESC and Stress values

Fig. **??** portrays temporal changes of stress values and ALs in facial expressions of Subject A. We next analyze effects between psychological stress and facial expressions with cor-

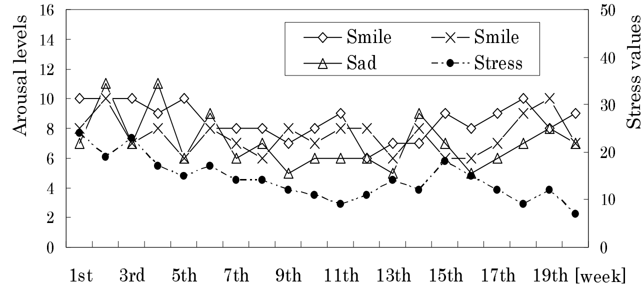


Fig. 7. Time-series changes of stress values and arousal levels (Subject A).

TABLE I
CORRELATION COEFFICIENT OF STRESS VALUES AND AROUSAL LEVELS.

Subject	Happiness	Angry	Sadness	FESC
A	**0.403**	-0.053	**0.308**	**0.306**
B	-0.007	-0.039	**-0.253**	**-0.241**
C	-0.190	-0.068	0.047	-0.183
D	0.174	-0.169	**0.732**	**0.445**
E	**0.526**	**0.584**	-0.183	**0.298**
F	**-0.527**	-0.093	**-0.213**	**-0.408**
G	**-0.510**	**0.389**	**-0.254**	**-0.229**
H	**0.380**	-0.155	**0.418**	**0.351**
I	**0.271**	**0.384**	**0.461**	**0.609**
J	-0.089	-0.138	-0.077	-0.163

relation coefficients. Table **??** portrays correlation coefficients between stress values and ALs in each facial expression of 10 subjects. Correlation coefficients show various patterns among subjects. Subject I is a typical case of a positive correlation. All facial expressions show positive correlations, especially for sadness, which shows 0.609. Subject F is a typical case of a negative correlation. Subjects C and J show no significant correlation coefficients for any facial expression. Particularly, the FESCs of Subject C have no marked changes, although she exhibits strong stress with a high value.

C. Estimation of stress levels

The degree of actual facial expressions is modified by various types of psychological effects, a situation, atmosphere, etc., although spontaneous and intentional facial expressions are triggered by emotional changes and intentional social

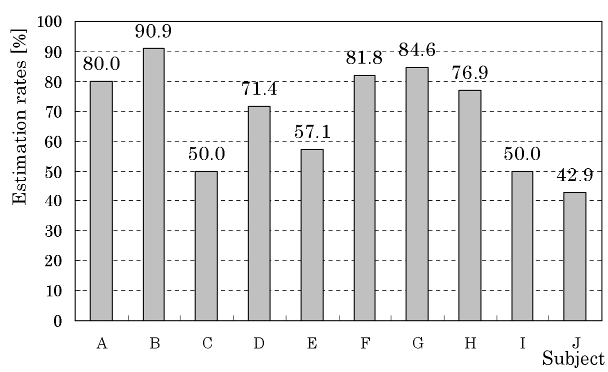


Fig. 8. Stress estimation results with SVM

restrictions, such as when one makes a fake smile. In this study, we have acquired facial expression images continually during a long period in an identical situation. The FESCs show various distributions in each week. Therefore, the ALs that show the degrees of expressions in our method differ each week. In this experiment, we specifically examine the effect between expressions and stress from psychology for estimating stress levels from FESCs.

We used SVMs [11], which have high recognition capability, for mapping input data to a high dimensional space using kernel tricks. We evaluated estimation rates using Leave-One-Out Cross Validation (LOOCV). The estimation targets are stress evaluation values of four steps: Level 1 (weak), Level 2 (normal), Level 3 (slightly high), and Level 4 (high). Herein, the distribution of target datasets is 48.6% of Level 2, 33.6% of Level 1, 13.1% of Level 3, and 4.7% of Level 1. We used Radial Basis Functions (RBFs) as a kernel function for SVMs. We set the parameter γ, which controls the distribution of Gauss functions, to the inverse number of input vectors.

Fig. 8 portrays stress estimation results of 10 subjects. The mean estimation rate is 68.6%. The highest estimation rate is 90.9% of Subject B. Subsequently, the estimation rates of Subjects G and F are, respectively, 84.6% and 81.8%. The estimation rate of Subject A, who had facial expression images taken the most times, is 80.0%. In contrast, the estimation rates of Subjects C and I are the same: 50.0%. The lowest estimation rate is 42.9% of Subject J. The data lengths of Subjects J and C are, respectively, only seven and eight weeks. We consider that this is a difficult problem for SVMs to create a classifier of four categories from such few data. Therefore, we selected the datasets of these subjects for more than 10 weeks. The mean estimation rate of Subjects A, B, F, G, H, and I is 77.4%. We consider that the estimation performance will be improved if long-term datasets of more than 10 weeks were obtained to continue to obtain vertical datasets.

Using our method, we achieved efficient estimation of stress levels, although we used SVMs under the condition of disproportionate training data distribution. We evaluated all datasets using LOOCV. The number for datasets for each stress level is various. The number for datasets of Level 2 is the largest: about 50%. This rate reaches 80% when including the number of datasets of Level 1. Moreover, five patterns of datasets were produced, which correspond to four subjects; one set of data consisted of stress levels. Six patterns of five subjects produced only two samples. We used all datasets of these few samples without exception, although it is difficult to learn and to estimate these samples using conventional generalization capabilities. To collect these data evenly is a challenging task because stress distributions vary among individuals. We consider that estimation performance will be improved to increase the terms of image acquisition, enabling us to address seasonal transformations.

V. Conclusion

This paper presents FESCs as a framework to describe individual facial expression spaces based on the consideration of facial expressions created by emotion as an individual space in each person. The ALs are created by categories that are classified by SOMs and integrated with Fuzzy ART. The FESCs are created with the axes of ALs of three facial expressions (happiness, anger, and sadness) based on the Russell's circumplex model. We created an original facial expression dataset of 10 subjects (five male subjects and 5 female subjects) for seven weeks. Using this dataset, our method can express individual facial expression spaces using FESCs. Moreover, we used SRS-18 for measuring the stress levels of each subject before taking images. We analyzed the effects of psychological stress using FESCs. The results show that happiness and sadness are affected by stress in most subjects.

Future studies must evaluate intentional and spontaneous facial expressions for discrimination using symmetry properties of the horizontal direction to represent facial expression rhythms created by individual patterns of time changes of ALs. Moreover, we will seek to increase the number of subjects for horizontal studies between subjects and to capture long-term datasets for vertical studies in each subject to analyze and to elucidate relations between facial expressions and stress.

Acknowledgment

The authors would like to thank fellow engineers at SmartDesign Co. Ltd. for providing much appreciated advice and for developing a system for taking facial images for these experiments. We also thank the 10 students at our university who participated as subjects by letting us take facial images over such a long period. 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, L. J.M. Rothkrantz, "Automatic Analysis of Facial Expressions: The State of the Art," IEEE Trans. PAMI, Vol.22, No.12, pp.1424-1445, Dec 2000.

[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] H. Madokoro, K. Sato, A. Kawasumi, and S. Kadowaki, "Facial Expression Spatial Charts for Representing of Dynamic Diversity of Facial Expressions," IEEE Int'l Conf. System, Man, and Cybernetics, (SMC), Oct. 2009.

[5] 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.

[6] G. Donato, M. S. Bartlett, J. C. Hager, P. Ekman, T. J. Sejnowski, "Classifying Facial Actions," IEEE Trans. PAMI, Vol.21, No.10, pp.974–989, Oct. 1999.

[7] T. Kanade, J.F. Cohn, and Y. Tian,"Comprehensive database for facial expression analysis," IEEE Face and Gesture, pp.46-53, 2000.

[8] M. Bartlett, J. Hager, P. Ekman, and T. Sejnowski. Measuring facial expressions by computer image analysis. Psychophysiology, 36:253-264, 1999.

[9] M. Pantic, M.F. Valstar, R. Rademaker and L. Maat, "Web-based Database for Facial Expression Analysis", Proc. IEEE Int'l Conf. Multmedia and Expo, Amsterdam, The Netherlands, July 2005

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

[11] V. Vapnik, Statistical Learning Theory, Wiley, 1998.