Comparative Study of Feature Extraction Method for Emotional Classification by Micro-expressions

Koki Kato*, Hironobu Takano*, Masahiro Saiko†, Masahiro Kubo† and Hitoshi Imaoka†
* Toyama Prefecutural University, Toyama, Japan
E-mail: takano@pu-toyama.ac.jp
† NEC, Kanagawa, Japan

***Abstract*—Human facial expressions include a slight and instantaneous movement called micro-expressions. Unlike ordinary facial expressions, micro-expressions are impossible to control by oneself. Since micro-expression shows true emotions, micro-expression recognition is expected to play an active role in clinical diagnosis and business negotiations. However, it is difficult to recognize micro-expression because of their insensible and quick facial movements. In this study, we aimed to improve the accuracy of emotional estimation using micro-expressions. In previous researches on emotional estimation using micro-expression, LBP-TOP (Local Binary Pattern from Three Orthogonal Planes) and CBP-TOP (Centralized Binary Patterns from Three Orthogonal Planes) have been utilized. However, it is unclear if the feature selection and the combination of multiple features for emotional classification are effective. In this study, the emotional classification was performed using selected components of each individual feature. In addition, we investigated whether the fusion of scores obtained from each feature improved the accuracy of emotional estimation. The experimental results showed that the accuracy of emotional classification was increased by feature selection, whereas the score level fusion did not contribute to improve the performance of emotional estimation.**

I. Introduction

In recent years, micro-expressions play an important role in emotional estimation. The micro-expression is very short, quick, and unintentional facial expressions[1]. It is difficult for human to recognize micro-expression which occurs for short duration and only involves a slight movement of a certain part of a face. To improve the ability to recognize micro-expressions, Ekman developed the METT (Micro-Expression Training Tool), which is a training tool to recognize micro-expressions classified into seven emotions. However, even with the training by METT, micro-expression recognition rate improved by only about 40%. Therefore, it is necessary to develop the micro-expression recognition method using machine learning. In previous studies, high-dimensional features have been used. Thus, the feature might have redundant description for emotional classification. In addition, it is unclear whether emotion classification method using multiple features provide an improvement of performance because only individual feature has been used in previous studies. In this study, we investigate the emotional estimation performance by micro-expressions using two types of features that are LBP-TOP and CBP-TOP. We evaluated the accuracy of emotional estimation for the score level fusion of the LBP-TOP and CBP-TOP. Furthermore, the feature selection by using the ratio of the inter-class variance to within-class one was employed. The effect on the feature selection was also discussed in this paper.

II. Related work

We describe emotional classification by micro-expressions and datasets of micro-expression.

A. Micro-expression datasets

Table I shows the micro-expression datasets: SMIC was created using three different types of cameras [2] ; CASME II is the largest and most widely used dataset [3] ; SAMM is the most culturally diverse dataset and various ethnic subjects participated for dataset construction [4] . SMIC has three types of emotions: Positive, Negative, and Surprise; CASME II has five types of emotions: Happiness, Surprise, Disgust, Repression, and Other; and SAMM has seven types of emotions: Happiness, Surprise, Disgust, Anger, Fear, Sadness, and Contempt. Micro-expressions can be measured by presenting a video that induces a certain emotion. The examiner instructs the subject to keep his or her facial expression as neutral while the subject is watching the video. The unintentional micro-expressions can be elicited by keeping facial expression as neutral.

TABLE I
Micro-expression datasets.

Datasets	Subjects	Samples	Classes of Emotion
SMIC (HS)[2]	16	157	3
SMIC (VIS)	8	71	3
SMIC (NIS)	8	71	3
CASME II[3]	26	247	5
SAMM[4]	32	159	7

B. Previous studies

Liong et al. selected ROI (Region of Interest) based on AU (Action Unit) for emotions and evaluated emotional estimation performance [5]. SMIC and CASME II were used as the datasets for the evaluation of emotional estimation. A feature, a classifier, and an evaluation method used for emotional estimation were LBP-TOP, SVM (Support Vector Machine), and LOVO (Leave One Video Out). The experimental result shows that the accuracy with the selected ROI is about 3%

better than that with the feature generated from the whole face.

Guo et al. used CBP-TOP as a feature extraction method and ELM (Extreme Learning Machine) as classifier to estimate emotions based on micro-expressions [6]. The LBP-TOP, which has been widely used for emotional estimation by micro-expressions, is a high-dimensional feature and has a problem of including noise. The CBP-TOP can be generated as 32-dimensional feature, while the LBP-TOP feature has 256 dimensional components. The accuracy of the CBP-TOP feature increased by about 8% compared to the LBP-TOP feature.

III. Proposed method

We describe the emotional estimation method proposed in this paper. First, the preprocessing and ROI selection for feature extraction are explained. Next, LBP-TOP and CBP-TOP features used for emotional classification are shown. Finally, we describe the feature selection method and classifier for emotional estimation.

A. Landmark detection of a face

This section describes the preprocessing required for feature extraction. Since the features used in this study are extracted from 2-dimensional images, the video frames in which the micro-expression appeared are extracted from original facial video. Face detection and landmark detection are performed on each video frame. The 68-point human face landmarks are plotted as shown in Fig. 1(a), and the actual detected landmarks are shown in Fig. 1(b). The plotted points consist of 17 points for the face contour and several points for each part of the face such as eyes and mouth. For landmark detection, we used the landmark detector included in Dlib [7].

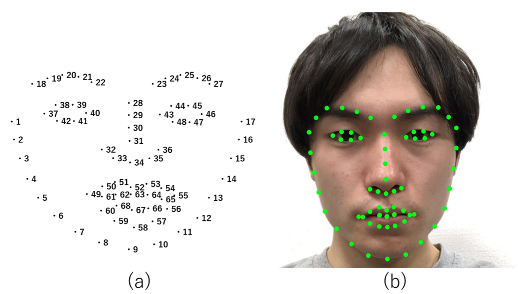

Fig. 1. Landmark detection of a face.

B. ROI selection

Since micro-expressions are defined as localized slight facial movements, the surface of the face hardly changes except for the regions related to the expression. Figure 2 shows the ROI selection method used in this study. The ROI can be selected using the landmarks described in the previous section. ROI-1 has three regions (right eye, left eye, and mouth), ROI-2 has two regions (glabella and mouth), and ROI-3 has two regions (both eyes including glabella and mouth).

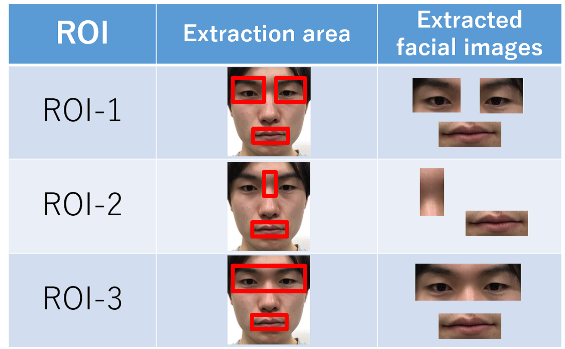

Fig. 2. ROI selection

C. LBP-TOP

The calculation method of the LBP-TOP feature is shown in Fig. 3. The LBP feature consists of relative values of each pixel obtained by comparing a center pixel with its surrounding ones. The LBP-TOP feature is an extension of the LBP feature which also includes temporal information. The LBP-TOP feature can be obtained by calculating LBP in the XY, XT, and YT planes and concatenating the histograms of each plane. Here, the XY plane represents a general image, and the XT and YT planes represent as each axis of the image plane and time axis.

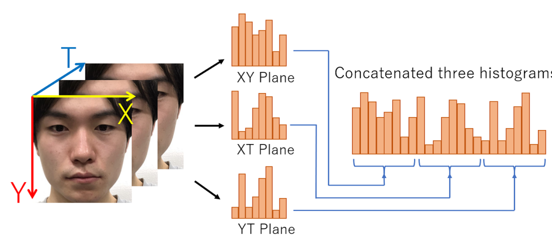

Fig. 3. LBP-TOP calculation method.

D. CBP-TOP

The calculation method of CBP features, which are the basis of CBP-TOP features, is shown in Fig. 4 [6]. The central pixel value is emphasized with the highest weight in the CBP feature extraction. While the LBP has 256 dimensions per plane and is high dimensional feature, the CBP can be generated as 32 dimensions per plane. The calculation method is described as follows. First, a color image is converted to a grayscale one. Next, the region of 3×3 pixels around the pixel of interest is extracted. For the region of interest, the difference between symmetrical pixel values around the pixel of interest is calculated and compared with an arbitrary threshold value. The center pixel value is compared with the average of the pixel values in the region of interest. The pixel value is set as 1 if the difference between the center pixel value and the average of the surrounding pixels is the same or larger than the threshold, otherwise the pixel value is set as 0. The pixel value is then interpreted as a binary number and converted to a decimal number as shown in Fig. 4. The calculation described above is performed for all pixels to obtain 32-dimensional feature values.

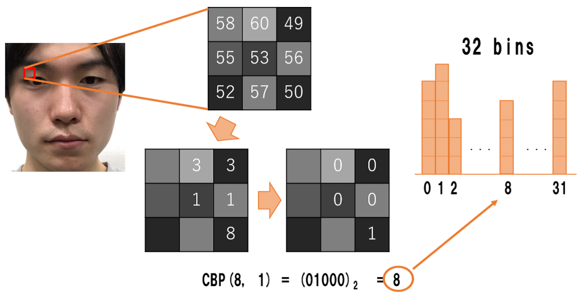

Fig. 4. CBP calculation method.

E. Feature selection

The feature selection is to extract effective component of feature vector from original one and to perform dimensionality reduction[8]. The ideal distribution of data for classification presents the condition in which data belonging to the same class are as close as possible and each class is significantly separately-placed. The degree of cohesion of data in the same class is shown by within-class variance, and the degree of spread among classes is presented by inter-class variance. The ratio of the inter-class variance to within-class one, which evaluates the degree of separation among classes, was used as an index for feature selection. Here, feature selection was performed for the histogram calculated in each plane. Based on the predetermined dimensionality reduction rate, the component of feature vector was selected in decreasing order of the ratio of the inter-class variance to within-class one. The reduction rates of 80%, 60%, 40%, and 20% were adopted in this study.

F. Ensamble learning

The XGBoost, one of the ensemble learning methods, was adopted for emotion classification [9]. The ensemble learning improves the prediction performance of unlearned data by combining multiple classifiers obtained by training differently. Boosting minimizes the loss function by adapting the weight of classifier that could not be discriminated by the previous classifiers, and uses the previous decision trees to train the next classifier. In addition, boosting can also provide the features selected at each node of the weak classifier and evaluate the reliability of estimation results.

IV. Expriments

A. Evaluation method

The performance evaluation of emotional classification was conducted by using accuracies of individual LBP-TOP and CBP-TOP, as well as the score level fusion of LBP-TOP and CBP-TOP, which are obtained from the whole face, ROI-1, ROI-2, and ROI-3 regions. The XGBoost was used as the classifier for emotional estimation, and score level fusion of LBP-TOP and CBP-TOP was conducted with predicted values of each feature. The fused score can be obtained by adding the weighted prediction values. The weights assigned to each score were arbitrarily set as 0.1 to 0.9 so that the total weight assigned to each score became 1. For the evaluation method, LOVO cross-validation was also used. This method treats one micro-expression video image in the dataset as test data and the rest as training data. In this analysis, the accuracy was calculated by repeating the training process so that all the data became the test data.

B. Experimental results

Figures 5 and 6 show the accuracy for LBP-TOP and CBP-TOP with and without feature selection in SMIC, and Table II shows the results for all datasets. The vertical axis is the accuracy and the horizontal axis is feature extraction regions. From the experimental results, the accuracy increases by selecting the features.

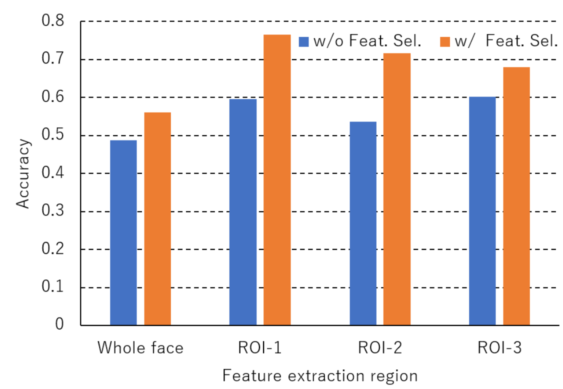

Fig. 5. Accuracy comparison with and without feature selection (LBP-TOP).

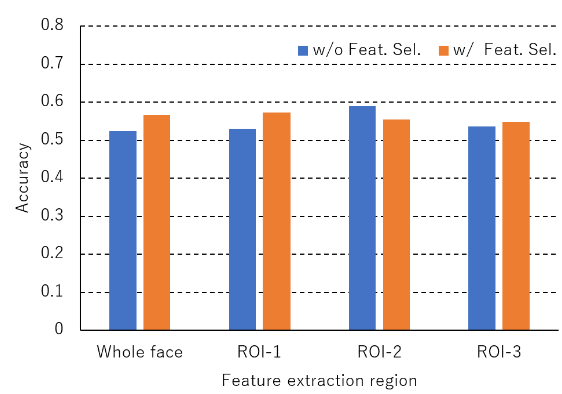

Fig. 6. Accuracy comparison with and without feature selection (CBP-TOP).

TABLE II
EMOTIONAL CLASSIFICATION ACCURACY WITH AND WITHOUT FEATURE SELECTION

		LBP-TOP				CBP-TOP			
		Whole face	ROI1	ROI2	ROI3	Whole face	ROI1	ROI2	ROI3
SMIC	w/o Feat. Sel.	0.488	0.596	0.536	0.602	0.524	0.530	0.590	0.536
	w/ Feat. Sel.	0.560	0.765	0.717	0.681	0.566	0.572	0.554	0.548
CASME II	w/o Feat. Sel.	0.474	0.502	0.522	0.522	0.498	0.570	0.522	0.534
	w/ Feat. Sel.	0.482	0.598	0.622	0.602	0.482	0.538	0.522	0.536
SAMM	w/o Feat. Sel.	0.534	0.504	0.511	0.466	0.444	0.474	0.481	0.481
	w/ Feat. Sel.	0.534	0.617	0.617	0.617	0.485	0.534	0.556	0.504

For each dataset, the accuracies for the individual feature and score level fusion are shown in Figs 7 – 9. These figures show the highest accuracies obtained by changing the feature selection rate of each feature or the weighted ratio for score level fusion. From the experimental results, the feature with the highest accuracy in all datasets was LBP-TOP. In particular, the accuracy of the LBP-TOP was significantly improved by the ROI selection. The accuracy of the score level fusion for the whole face of SMIC and CASME II was better than that of the respective features. However, in the other results, the accuracy of score level fusion did not exceed one of the LBP-TOP.

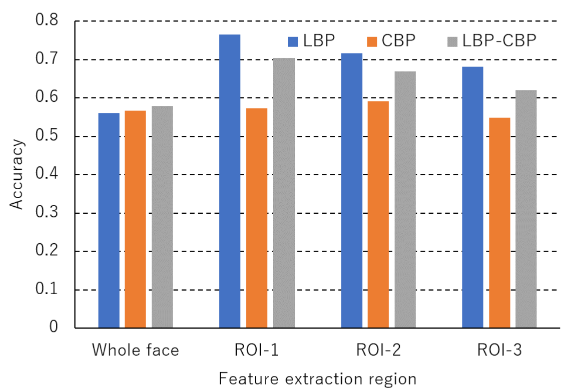

Fig. 7. Emotional classification accuracy for each feature (SMIC).

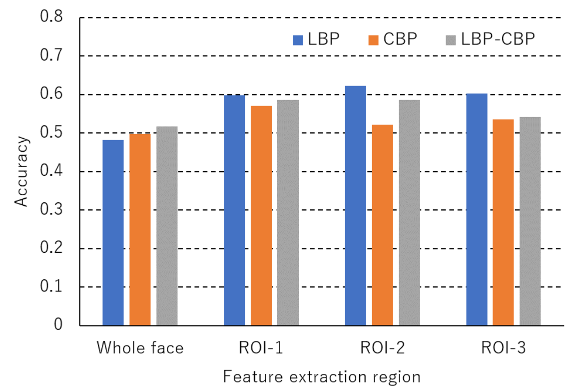

Fig. 8. Emotional classification accuracy for each feature (CASME II).

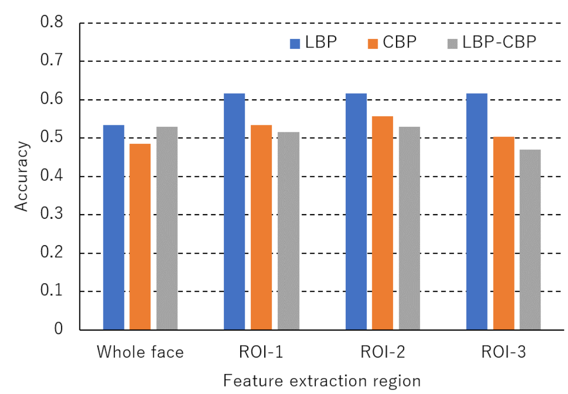

Fig. 9. Emotional classification accuracy for each feature (SAMM).

C. Discussion

We discuss in terms of feature selection, ROI selection, and score fusion. In all datasets, the feature with the highest accuracy was the feature-selected LBP-TOP feature. Thus, the LBP-TOP is the optimum feature for emotional estimation by micro-expressions under the condition of ROI selected in this study. Feature selection by the ratio of the inter-class variance to within-class one is an effective method of improving the accuracy of emotion estimation using micro-expressions. In selection of ROI, the ROIs with the highest accuracy for each dataset were different. The ROI-1, which was selected in the previous study, showed the highest accuracy in SMIC. However, for the rest of datasets, there was no significant difference among ROIs. Thus, the suitable ROIs might depend on the number of emotion classes or the types of emotions for classification. Since the accuracy was not necessarily improved by score level fusion, LBP-TOP and CBP-TOP include similar characteristics for emotional classification.

V. CONCLUSION

In this study, we used the feature-selected LBP-TOP, CBP-TOP, and the fused scores of these features in emotional estimation using micro-expressions. The highest accuracy in all datasets was feature selected LBP-TOP. The feature selection for LBP-TOP and CBP-TOP increases the accuracy of emotional classification. In addition, for the score level fusion, the emotional estimation performance tends to be affected by the feature with low score. Since the suitable feature depends on the type of emotion and ROI, it is necessary to clarify the partial face regions in which the effective feature for emotional estimation using micro-expressions are extracted. We also plan to investigate the emotional classification performance with the HOG-like features including gradient information of pixel values.

References

[1] P. Ekman and W. V. Friesen, Facial Action Coding System: A technique for the measurement movement, Consulting Psychologists Press, 1978.
[2] X. Li, T. Pfister, X. Huang, G. Zhao, and M. Pietikainen, “A Spontaneous Micro-expression Database: Inducement, Collection and Baseline”, Proc. 2013 10th IEEE International Conference and Workshops on Automatic Face and Gesture Recognition, pp.1–6, 2013.
[3] W.-J. Yan, X. Li, S.-J. Wang, G. Zhao, Y.-J. Liu, Y.-H. Chen, and X. Fu, “CASME II: an improved spontaneous micro-expression database and the baseline evaluation”, Plos One, Vol.9, No.1, pp.1–8, 2014.
[4] A. Davison, C. Lansley, N. Costen, K. Tan, and M. H. Yap , “SAMM: a spontaneous micro-facial movement dataset”, IEEE Trans. Affect. Comput., Vol.9, No.1, pp.116–129, 2016.
[5] S. Liong, J. See, R. Phan, K. Wong, and S. Tan, “Hybrid facial regions extraction for micro-expression recognition system”, J. Sign. Process. Syst., Vol.90, pp.601–617, 2018.
[6] Y. Guo, C. Xue, Y. Wang, M. Yu, “Micro-expression recognition based on CBP-TOP feature with ELM”, Optik, Vol.126, No.4, pp.4446–4451, 2015.
[7] dlib C++ Library, http://dlib.net
[8] R. O. Duda, P. E. Hart, and D. G. Stork, Pattern Classification Second Edition, John Wiley & Sons, 2001.
[9] T. Chen and C. Guestrin, “XGBoost: A scalable tree boosting system”, Proc. the 22nd ACH SIGKDD Int. Conf. Knowledge Discovery and Data Hining, pp.785–794, 2016.