Person-invariant Recognition of Subtle Smiles using Selected Improved LBP Features

Kosuke SASAKI and Manabu HASHIMOTO
Graduate School of Computer and Cognitive Sciences
Chukyo University
Aichi, Japan
Email: {sasaki, mana}@isl.sist.chukyo-u.ac.jp

Noriko NAGATA
Graduate School of Science and Technology
Kwansei Gakuin University
Hyogo, Japan
Email: nagata@kwansei.ac.jp

***Abstract*—This paper describes a person-invariant method for classifying subtle smile. The system uses keypoints detected by using a face tracking tool called Face Tracker. Our method selects features strategically by calculating their human dependency rate. Its classification accuracy was evaluated using the facial images of unlearned people. The results showed the average F-measure was 0.87 for neutral (expressionless) facial images, 0.78 for subtle smile images, and 0.87 for exaggerated smile images. Also, person-invariant accuracy was evaluated by using the F-measure frequency of unlearned people. The results revealed that the proposed method has higher person-invariant accuracy than the previously reported methods.**

Keywords—facial keypoints; person-invariant recognition; Improved LBP

I. INTRODUCTION

Recently, many people have become interested in improving the quality of human life. Because people are always looking for ways to improve their quality of life, product makers and service providers desire a system that can estimate human emotions so they can develop products and services towards with customers will feel positively. One study for estimating human emotions reported a method using an electroencephalogram (EEG) as a contact sensor [1]. However, the sensor is uncomfortable to wear, and this disturbs the expression of natural emotion. To avoid such disturbance, human emotion needs to be estimated in a natural situation. Since facial expressions are closely related to emotions, many researchers have studied facial expressions by acquiring them by using a non-contact sensor such as an RGB (red, green, and blue) camera. Two types of methods for classifying facial expressions have previously been reported. One uses local facial features such as HoG [2] and Gabor wavelets [4]. The other uses a small set of keypoints detected from parts of the face [5] [6]. These methods can classify many facial expressions but cannot recognize very subtle ones. Consequently, Matsuhisa and Hashimoto [7] proposed a method to recognize subtle smile by using Gabor filters and the AdaBoost algorithm. Gabor filters have a response value that is sensitive to subtle changes in smile. However, they have difficulty classifying smile of unlearned people because the response value represents a 3D facial shape that is different for each person. Also, Nomiya and Hochin [8] proposed a method to recognize subtle facial expressions by using the geometric features of keypoints. However, this method has the same problem as Matsuhisa and Hashimoto's method because the placement of the facial parts is different for each individual. Local binary pattern features has person-independent characteristics [3]. In particular, Faisal et al. [9] proposed a person-independent method to recognize facial expressions by using compound local binary patterns. However, this method cannot recognize subtle facial expressions because it selects only features that change significantly. Thus, previous methods have two problems. First, they cannot recognize subtle facial expressions. Second, identification performance is dependent on using learned people as subjects. In particular, we focus on smile in facial expressions because the smile is closely related with positive emotion such as happiness and satisfaction. To address these problems, we propose a person-independent method for classifying smile intensity. In the proposed method, keypoints are extracted by using a face tracking tool called Face Tracker [10]. Improved local binary pattern (LBP) features [11] are calculated by comparing the pixel values of a keypoint with the pixel values around the keypoint. Person-invariant features are selected by using the human dependency rates of calculated features. Input facial images are classified by using weights as human dependency rates. These factors enable the proposed method to recognize subtle smile of unlearned people.

II. PROPOSED METHOD FOR CLASSIFYING SMILE INTENSITY

A. Overview

Figure 1 shows a block diagram of the method we propose using to classify smile intensity.

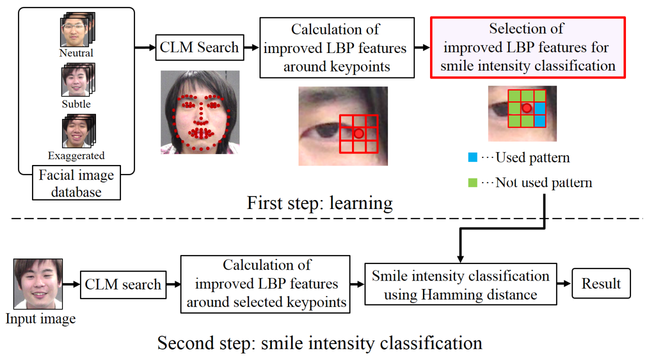


Fig. 1. A block diagram of the proposed method.

The proposed method comprises two steps, the first for learning and the second for classifying. In the first step, facial keypoints from a local position such as a corners of the eyes or a mouth corner are detected by using a constrained local model (CLM) method that is versatile to humans. Figure 2 shows detected keypoints obtained from facial images.

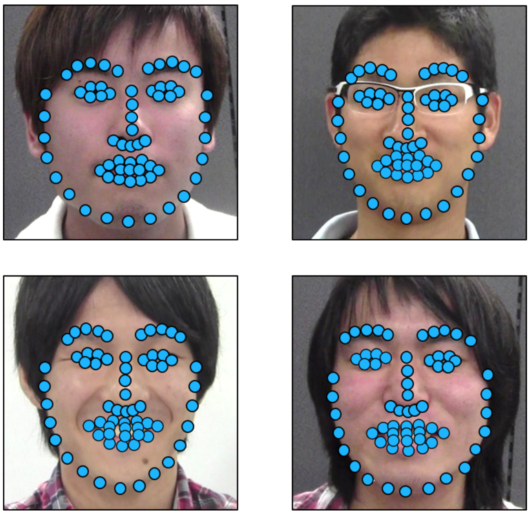


Fig. 2. Detected keypoints from facial images.

This method detects 66 keypoints in all. However, the 17 keypoints obtained for the facial outline are unusable as background information, so we used only 49 keypoints. In addition, the size and rotation of a facial image must be corrected to calculate the features stably. The selected keypoints are corrected by using the relationships among keypoints such as the horizontality of the centers of both eyes and the center of the gravity point that comprises the centers of both eyes and the upper lip. Improved LBP features that are represented as 0 or 1 bits are calculated by comparing the keypoint pixel values with the values around the keypoint. The improved LBP features can capture the presence or absence of edges better than ordinary LBP features. Therefore, they can capture the presence or absence of facial changes. The features are person-invariant because the magnitude of pixel value variation is independent. In addition, the regions around the facial keypoints are slightly binary. The human dependency rates are calculated in each region; our method selects the region having the lowest dependency rate. Bit streams are generated in smile intensity using the selected regions. In addition, features are selected for evaluating identities on the basis of smile intensity. The most effective bits for classifying smile intensity are selected from bit streams having low human dependency rates. The selected features are utilized to classify smile intensity on the basis of two requirements. First, the features must be human-independent. Second, they must be effective for classifying smile intensity. In second step, the keypoints are detected from input images in the same way as in first step. The improved LBP features are calculated in each keypoint. Further, the bit streams are generated by using the regions selected in first step. Our method classifies the smile intensity of input images by computing the Hamming distance between the bit streams of first step and second step.

B. *Improved Local Binary Patterns[11]*

The LBP features express the spatial structures of the local textures of gray scale images by thresholding pixel neighborhoods with the center values. Because these features capture the brightness difference between focused pixels and the area around the pixels, they can capture the presence or absence of brightness around keypoint edges. However, they cannot capture subtle edges. Thus, improved LBP features were previously reported by Hongliang et al. [11]. These improved LBP features can capture subtle edges better than ordinary LBP features can. Figure 3 shows the flow of computing ordinary LBP features and the improved LBP features.

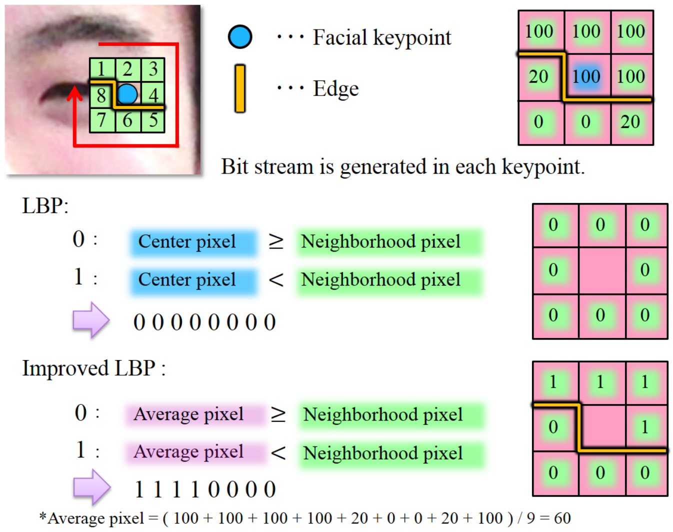


Fig. 3. Computing LBP features and improved LBP features.

The improved LBP features use the average of neighborhood pixel values. Thus, we can confirm that these features have a local structure, which is obtained successfully as shown in Figure 3. Our method uses not only a 3x3 neighborhood region but also 5x5, 7x7, 9x9 and 11x11 neighborhood regions. Thereby, with it we can capture not only subtle but also substantial facial changes.

C. *Feature selection based on human dependency rate and facial expression identities*

Parts of the improved LBP features have personal characteristics such as wrinkles and sculpted features. We calculated the human dependency rates from bit frequencies in a learning dataset. Figure 4 shows the flow of calculating human dependency rates.

Our method focuses on the neighborhood regions around keypoints. We calculated the frequency values of binary bits of the regions that were focused on in the learning dataset; the dependency rate is an inverse value that is the difference between frequency values of 0 and 1. The person-independent regions are selected by thresholding this dependency rate. Moreover, the selected regions are given to the dependency rate as weights. The regions are also selected on the basis of facial expression identities. Our method selects effective regions to classify facial expressions in each class (neutral face

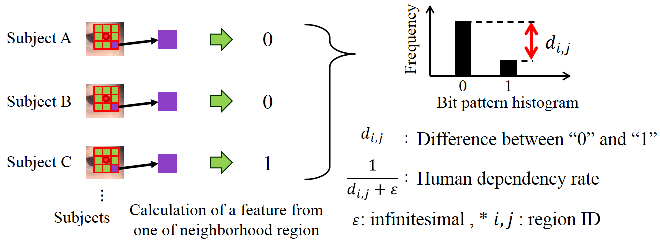


Fig. 4. Flow of calculating dependency rate on human.

and subtle smile, subtle smile and exaggerated smile, neutral face and exaggerated). These regions must differ slightly from one facial expression to another. On this basis, the bit streams that utilize facial expression classification are selected on the basis of two requirements. First, they must have low human dependency rate. Second, they must be identical for each facial expression.

D. Flow of Smile Intensity Classification

Figure 5 shows flow of smile intensity classification.

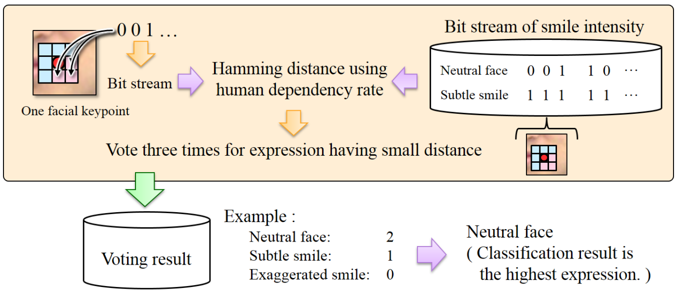


Fig. 5. Flow of smile intensity classification.

The proposed method detects facial keypoints from input images in the same way as learning images and calculates improved LBP features from keypoints of the input images selected by first step. The bit streams are generated from bits of the same regions as those selected in first step. Input images are classified by using the Hamming distance between the bit streams of input images and the bit streams generated in first step. This Hamming distance is calculated by using weights that consist of bits having human dependency rates. Because there are three classifiers, our method classified the input images by the voting of the classifiers.

III. Experimental Results

A. Dataset

We videoed 28 men and women watching a comedy program and used the captured images in an experiment. Five other people watched the taken video at the normal frame rate and typed in the ground truth to the video. They pressed 1, 2, or 3 when a someone made a neutral face, a subtle smile, or an exaggerated smile, respectively. We used the images given the same ground truth by at least four of the five people who watched the video. Figure 6 shows example images of each facial expression in the experiment.

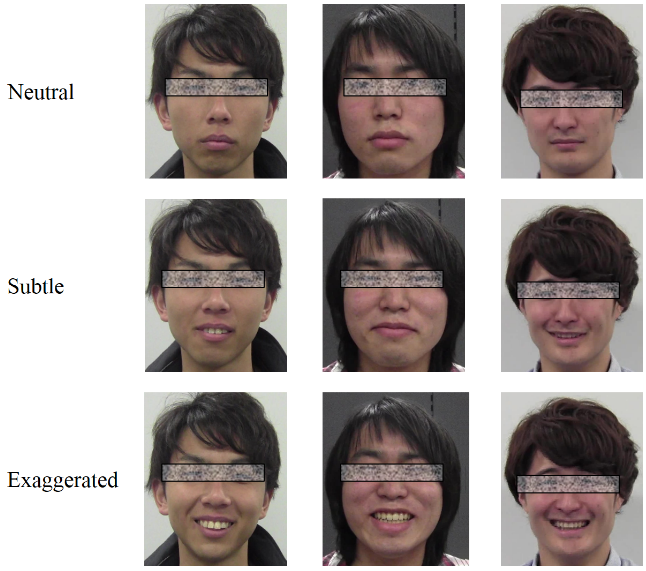


Fig. 6. Example images of each facial expression.

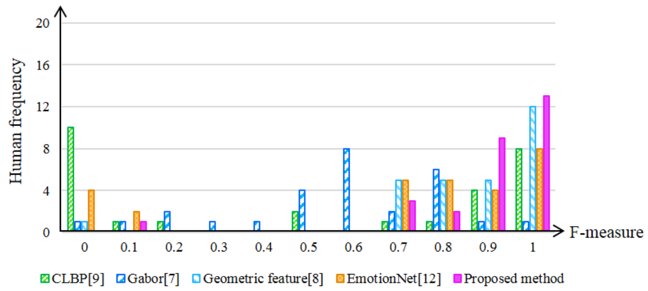
B. Smile intensity classification performance for unlearned people

By using a cross-validation approach, a classifier was trained by 27 people, and unlearned people were classified by it. Using 30 images for each facial expression, we also compared our method with those reported elsewhere [7] [8] [9] [12]. Faisal et al. [9] reported a facial expression classification method that considered person independence. Matsuhisa and Hashimoto [7] and Nomiya and Hochin [8] reported a recognition method for subtle facial expressions. Bradley and Adam [12] proposed EmotionNet that is a method of Convolutional Neural Networks (CNN) to emotion recognition in facial image. We utilized a learning CNN model that generated by the 15000 iterations. Figure 7 shows the F-measure frequency of unlearned people.

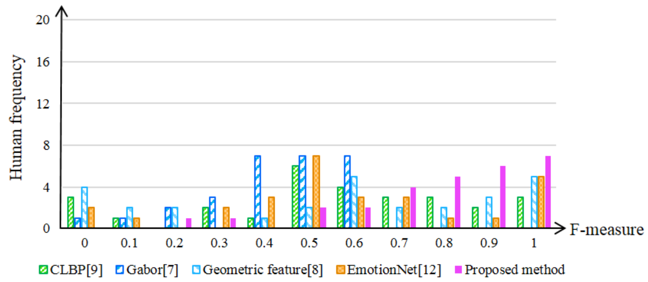
This graph shows the classification accuracy becomes person-independent as the distribution is concentrated on the right side. We confirmed that the results of the proposed method were distributed more to the right than those of previous methods. This shows that the proposed method is more person-invariant than the previous methods. In addition, the median value of frequency is calculated for each facial expression. Table 1 shows the median value of frequency for each facial expression.

TABLE I. Median Value of Frequency for each Facial Expression

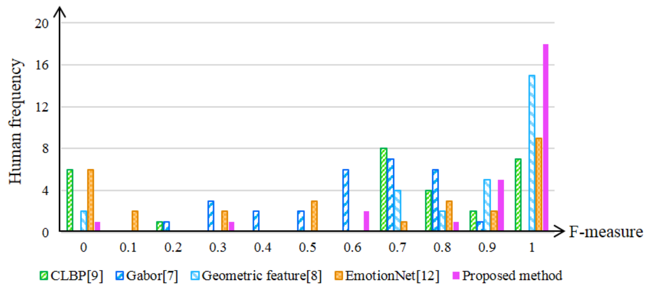
	The median value of frequency in Fig.7.		
	Neutral	**Subtle**	**Exaggerated**
Proposed method	0.9	0.8	1.0
EmotionNet [12]	0.8	0.5	0.7
Geometric feature [8]	0.9	0.6	1.0
Gabor filter [7]	0.6	0.4	0.6
CLBP [9]	0.5	0.6	0.7



(a) Frequency of F-measure value in neutral face



(b) Frequency of F-measure value in subtle smile



(c) Frequency of F-measure value in exaggerated smile

Fig. 7. F-measure frequency of unlearned people.

Our method provides similar or higher person-independence than previous method in neutral face and exaggerated smile. Our method provides higher person-independence than previous method in subtle smile.

We evaluated the classification accuracy for each facial expression. Table 2 shows the average F-measure for each facial expression.

TABLE II. F-MEASURE FOR EACH FACIAL EXPRESSION

	F-measure value		
	Neutral	**Subtle**	**Exaggerated**
Proposed method	0.87	0.78	0.87
EmotionNet [12]	0.68	0.57	0.57
Geometric feature [8]	0.85	0.57	0.85
Gabor filter [7]	0.58	0.43	0.61
CLBP [9]	0.51	0.56	0.63

The results showed the average F-measure was 0.87 for neutral images, 0.78 for subtle smile images, and 0.87 for exaggerated smile images. This confirmed that our method provides higher classification performance than previous methods.

IV. CONCLUSION

In this paper, we proposed the idea for achieving person-invariant classification of smile intensity. The idea was to select features by taking person invariance and facial expression identities into account. Person-invariant accuracy was evaluated by using the F-measure frequency values of unlearned people. For unlearned people, the values were 0.87 for neutral (expressionless) facial images, 0.78 for subtle smile images, and 0.87 for exaggerated smile images. We also found that these values were distributed at higher frequencies than the F-measures of previous methods. These results lead us to conclude that the proposed method can accurately recognize subtle smile and is person-independent.

ACKNOWLEDGMENT

This research is partially supported by the Center of Innovation Program from Japan Science and Technology Agency, JST.

REFERENCES

[1] C.A. Kothe, S. Makeig, and J.A. Onton, Emotion Recognition from EEG During Self-Paced Emotional Imagery, in Humaine Association Conference on Affective Computing and Intelligent Interaction (ACII2013), pp. 855-858, 2013.

[2] Junkai Chen, Zenghai Chen, Zheru Chi, and Hong Fu, Facial Expression Recognition Based on Facial Components Detection and HOG Features, Scientific Cooperations International Workshops on Electrical and Computer Engineering Subfields, pp. 64-69, 2014.

[3] C. Shan, S. Gong, and P. McOwan, Facial Expression Recognition Based on Local Binary Patterns: A Comprehensive Study, Image and Vision Computing, vol. 27, no. 6, pp. 803-816, 2009.

[4] E. Owusu, Y. Zhan, and Q.R. Mao, A neural-AdaBoost based facial expression recognition system, Expect Syst. Appl. 41(7), pp. 3383-3390, 2014.

[5] A. Majumder, L. Behera, and V. K. Subramanian, Emotion recognition from geometric facial features using self-organizing map, Pattern Recognition. 47(3), pp. 1282-1293, 2014.

[6] I. Kotsia, and I. Pitas, Facial expression recognition in image sequences using geometric deformation features and support vector machines, IEEE Trans. Image Process., vol. 16, pp. 172–187, 2007.

[7] H. Matsuhisa and M. Hashimoto, Identifying Subtle Facial Expression Changes using Optimized Gabor Features, The Journal of the Institute of Image Information and Television Engineers, vol. 68, no. 6, pp. J252-J255, 2014. (In Japanese)

[8] H. Nomiya and T. Hochin, Efficient Emotional Video Scene Detection Based on Ensemble Learning, The IEICE Transactions on Information and Systems, vol. J95-D, no. 2, pp. 193-205, 2012. (In Japanese)

[9] Faisal Ahmed, Hossain Bari, and Emam Hossain, Person-Independent Facial Expression Recognition Based on Compound Local Binary Pattern (CLBP), IAJIT, Vol.11, No, 2, pp.195-203, 2014.

[10] J. Saragih, S. Lucey, and J. Cohn, Deformable Model Fitting by Regularized Landmark Mean-Shifts, International Journal of Computer Vision, vol. 91, no. 1, pp. 200-215, 2011.

[11] Hongliang Jin, Qingshan Liu, Hanqing Lu, and Xiaofeng Tong, Face Detection Using Improved LBP Under Bayesian Framework, In Proceedings of Third International Conference on Image and Graphics, pp. 306–309, 2004.

[12] Bradley Kennedy and Adam Balint, EmotionNet, github: https://github.com/co60ca/EmotionNet