Person-Independent Classification of Subtle Facial Expressions using “Movement Direction Code” of Keypoints

Kosuke Sasaki, Ryota Watanabe, Manabu Hashimoto
Graduate School of Information Science and Technology
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-independent method of classifying subtle facial expressions. The method uses keypoints detected by using a face tracking tool called “Face Tracker”. It describes features such as coded movements of keypoints and uses them for classification. Its classification accuracy was evaluated using the facial images of unlearned people. The results showed the average F-measure was 0.88 for neutral (expressionless) facial images, 0.80 for subtle smile images, and 0.85 for exaggerated smile images.**

Keywords—facial expression classification; facial keypoint; person-independent;

I. Introduction

Recent years have seen considerable progress made in human-system coexistence, and it is expected that this will improve the quality of life in society. Towards this end, it will be necessary to use systems to estimate human emotions.

One study for estimating human emotions reported a method using a contact sensor like an EEG sensor [1]. However, people feel uncomfortable to wear and this disturbs the expression of natural emotions with sensor. Since facial expressions are closely related to emotions, it is important to classify them. Facial expressions can be acquired by using a non-contact sensor like an RGB camera.

Two types of methods for classifying facial expressions have previously been reported. One uses local facial features such as local binary patterns [2] and Gabor wavelets [3]. The other uses a small set of keypoints detected from parts of the face [4] [5]. These methods can classify many facial expressions, but they cannot recognize very subtle ones.

Consequently, Matsuhisa et al. [6] proposed a method to recognize subtle facial expressions with Gabor filters and the AdaBoost algorithm. Gabor filters have a response value that is sensitive to subtle changes in facial expression. However, it is difficult for them to classify facial expressions of unlearned people because the response value represents a 3D facial shape that is different for each person.

Also, Nomiya et al. [7] proposed a method to recognize subtle facial expressions by using the geometric features of keypoints. However, this method has the same problem as Matsuhisa’s because the placement of the facial parts is different for each individual.

Thus, there are two common problems with previous methods. First, they cannot recognize subtle facial expressions. Second, identification performance is dependent on using learned people as subjects. To address these problems, we propose a person-independent method of classifying subtle facial expressions.

Keypoints are extracted by using a face tracking tool called “Face Tracker” [9]. The radius and angles of moving keypoints are calculated and quantized from neutral (expressionless) facial images to subtle facial expression images or exaggerated facial expression images. The radiuses and angles are also coded. We define them as “Movement Direction Code”. In addition, keypoints based on the following two requirements are selected.

1. Difference of the “Movement Direction Code” is slight for each person.
2. Difference of the “Movement Direction Code” is great for each facial expression.

These factors enable the method to recognize subtle facial expressions of unlearned people.

II. Basic Idea

A. Background

Ekman et al. [8] reported that primary emotions correspond to typical facial expressions, and typical facial expressions are common to human beings. When people smile, for example, the mouth corners move backward and the lower lip moves downward. Such changes are also common to human beings. We define such changes as features.

B. Analyzing radius and angle similarities among moving keypoints

We examined whether the moving keypoint radius and angle were very similar from person to person when facial expressions change. The radius and angle were calculated for the left mouth corner for three test subjects. For the angle, the difference from person to person was less than ±10 degrees and the error was less than ±3 degrees. This confirmed that the angle range was very similar for each person.

However, the radius was found to be different for each person. Accordingly, we normalized all the radiuses on the basis of maximum radius. Using the normalized radius, which takes a value from 0 to 1, we found the difference from person to person was ±0.05 at maximum and the average difference among facial expressions was 0.3. This confirmed that there was a small radius difference from person to person but a large difference from facial expression to facial expression.

The obtained results confirmed that the moving keypoint radiuses and angles are very similar from person to person. Accordingly, we propose a feature that uses the radiuses and angles. The feature has two characteristics:

1. It is highly versatile from person to person.
2. It effectively classifies facial expressions.

III. PROPOSED METHOD

A. Overview

Figure 1 shows a block diagram of the method we propose to classify facial expressions.

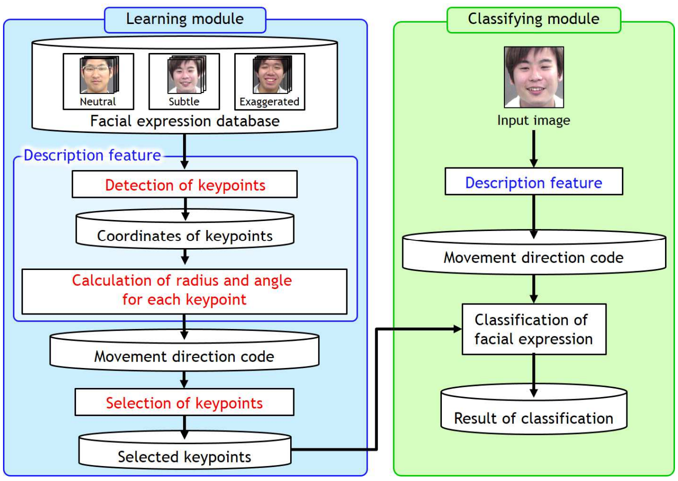

Fig. 1. Block diagram of proposed method.

The proposed method comprises two modules, module 1 for learning and module 2 for classifying. Module 1 detects keypoints from a local position in the face by using the CLM method [9]. This method is highly flexible from person to person and detects 66 keypoints in all. We used only 49 of them because the background makes the keypoints of the facial outline unusable. We also used the geometric relationship among keypoints to correct the size and rotation of facial images. The facial image is converted into the front direction.

We calculated and quantized the moving keypoint radiuses and angles, ranging from those for neutral facial images to those for expressive ones. This enhances the commonality of the radiuses and angles. We define this as a feature called "Movement Direction Code", which is calculated from all keypoints. Next, the feature is calculated from learning data and its occurrence frequency is generated for each facial expression. The feature is selected on the basis of two requirements: first, it must be highly similar for each person, and second, it must differ from one facial expression to another.

In module 2, keypoints are detected in the same way as in module 1. The feature is calculated by using the radius and angle of moving keypoints. For each facial expression, the feature's occurrence probability is calculated in module 1 and used to classify facial expressions.

B. Movement Direction Code

The "Movement Direction Code" feature enhances the person-independent characteristic because it quantizes moving keypoint angle and radius from neutral to expressive facial images. An example of the feature is shown in Fig. 2.

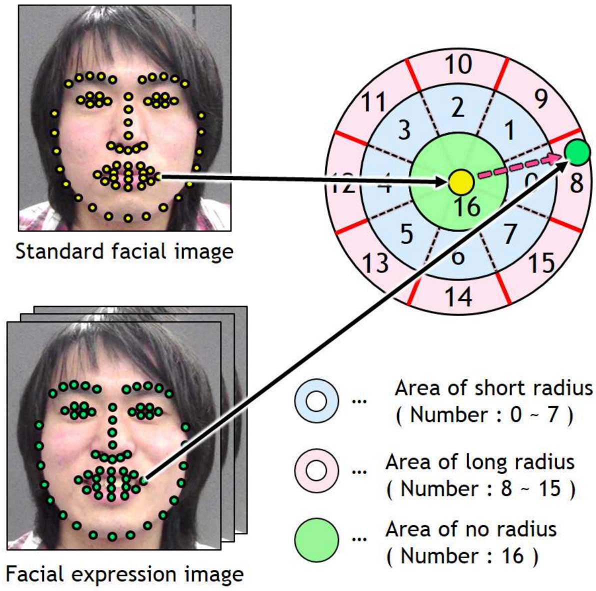

Fig. 2. "Movement Direction Code" example.

We calculated the radius and angle for each keypoint, which include keypoints of both standard facial images and facial expression images. A raw data of the angle causes the difference between individual. Since the raw data of the radius is the same for each person, each angle and radius is quantized. We call it the "Movement Direction Code" and define it as a feature for classifying facial expressions. Each of its codes has an occurrence probability of radius and angle for each facial expression. The occurrence probability is used to classify facial expressions. Keypoints are selected on the basis of two viewpoints:

1. Difference of the "Movement Direction Code" is slight for each person.
2. Difference of the "Movement Direction Code" is great for each facial expression.

Only keypoints that satisfy both viewpoints are used to classify facial expressions.

C. Automatic determination of angle resolution

Since the angle difference varies from person to person in each keypoint, it is necessary to determine the quantifying level number of angle in each keypoint. Figure 3 shows the flow of the automatic decision process for angle resolving.

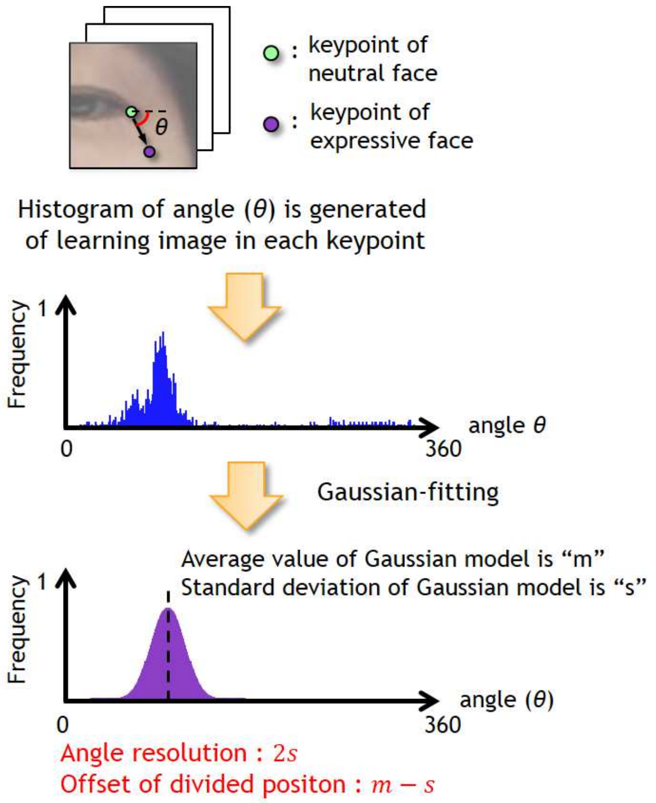

Fig. 3. Automatic angle resolving decision process.

The keypoint angle from a neutral face image to an expressive face image in learning data is calculated and the angle histogram is generated. A Gaussian model having an average value (m = 0, 1,..., 359) and standard deviation (s = 0.00, 0.05, 0.10,..., 25.00) is approximated and the model for which the sum of the errors is the smallest in each bin is selected. The angle resolution is determined by the standard deviation range of the selected model. The standard deviation and the average value are applied as an offset of the divided position for the "Movement Direction Code".

D. Dividing the facial keypoint radius to recognize subtle facial expressions

To recognize subtle facial expressions, it is necessary to determine a radius threshold that is similar for each person. However, since the radius is different for each person, we normalized all radiuses on the basis of maximum radius. The normalized radius, which take a value from 0 to 1, are used to calculate the radius thresholds r1 and r2. The r1 radius separates neutral facial expressions from those showing subtle smiles. The r2 radius separates facial expressions showing subtle smiles from those showing exaggerated smiles. Figure 4 shows decided threshold of radius using discriminant analysis method.

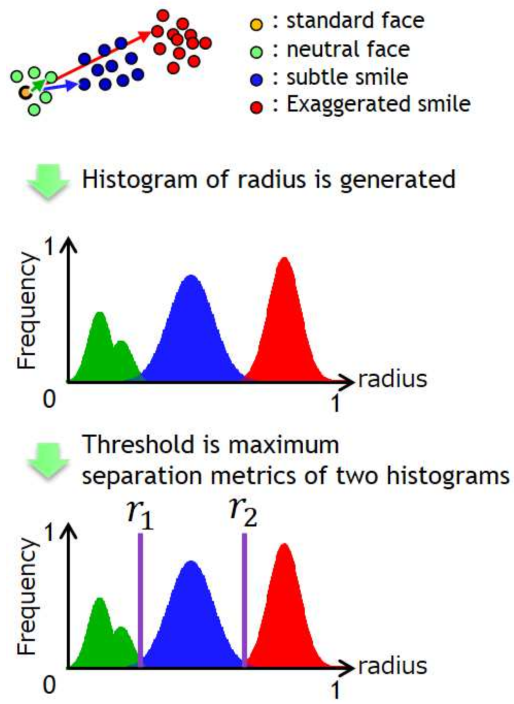

Fig. 4. Decided threshold of radius using discriminant analysis method.

The thresholds are calculated by using a discriminant analysis method. In this method, the threshold is maximum separation metrics of two histogram. The separation metrics is calculated from between-class variance and within-class variance.

E. Classification using facial expression occurrence probability

Figure 5 shows a flow of classification using occurrence probability.

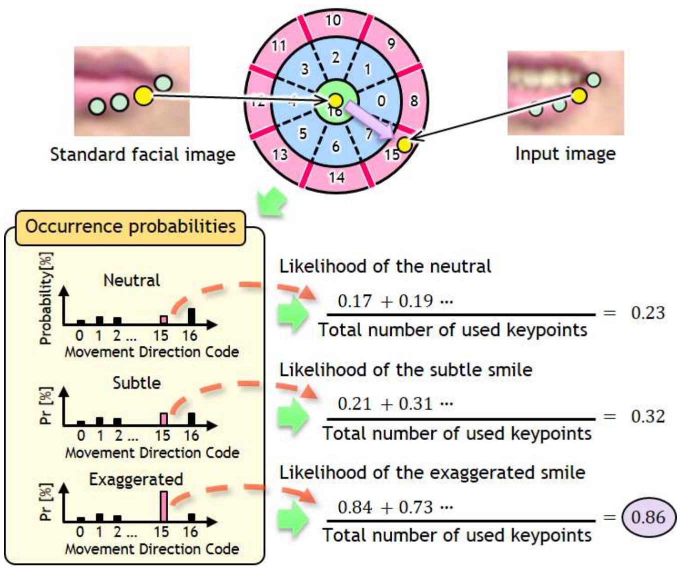

Fig. 5. Flow of classification using occurrence probability.

Facial expressions in input images are classified by using the feature's occurrence probabilities. These probabilities are calculated in the same way as in the learning module. Since they are able to represent the likelihood of facial expressions occurring, their average values for selected keypoints are calculated for each facial expression. The classification results obtained show the expressions having the highest occurrence probability on average.

IV. Experimental Results

A. Dataset

We took a video of 17 persons watching a comedy TV program and used the captured images in an experiment. Each of the captured facial expressions was labeled manually. Figure 6 shows example pictures of each facial expression used in the experiment.

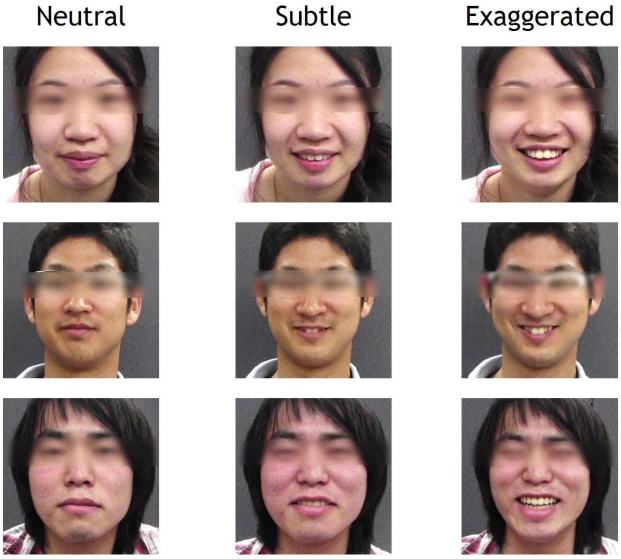

Fig. 6. Example pictures of each facial expression.

B. Expression classification performance for learned people

We used five learning images and 20 test images to classify the facial expressions of five learned people. Table 1 shows the F-measure obtained for each person.

TABLE I. F-measure for each person

	Neutral	Subtle	Exaggerated
Subject A	0.89	0.82	0.93
Subject B	0.95	0.80	0.83
Subject C	0.81	0.81	0.83
Subject D	1.0	1.0	1.0
Subject E	0.86	0.81	0.88

The results showed the average F-measure was 0.90 for neutral images, 0.86 for subtle smile images, and 0.89 for exaggerated smile images. Thus, for learned people we confirmed that the F-measure was higher than 0.80 for each facial expression.

C. Expression classification performance for unlearned people

Through the use of a 17-fold cross-validation approach, a classifier was trained by 17 persons and unlearned persons were classified by it. Using 25 images for each facial expression, we also compared our method with those reported in [6] and [7]. Table 2 shows the average F-measure obtained for each facial expression with the three methods.

TABLE II. Experimental Results For Each Facial Expression

	Neutral	Subtle	Exaggerated
Proposed method	0.88	0.80	0.85
Geometric feature [7] + AdaBoost	0.69	0.13	0.64
Gabor feature + AdaBoost [6]	0	0.38	0.43

The results showed the average F-measure was 0.88 for neutral images, 0.80 for subtle smile images, and 0.85 for exaggerated smile images. These results are similar to those shown in Table 1 and thus confirm that our method provides highly flexible identification performance from person to person. Therefore, we can conclude that the proposed method is person-independent.

V. Conclusion

We used three ideas to develop a method that achieves person-independent classification of subtle facial expressions. The first idea was to use features representing movements of facial parts. The second was to select features that differed only slightly for each person. The third was to select features that differed greatly for each person. With the method the average F-measure for learned people was found to be 0.90 for neutral images, 0.85 for subtle smile images, and 0.89 for exaggerated smile images. For unlearned people the respective values were 0.88, 0.80, and 0.85. These results lead us to conclude that the proposed method is able to accurately classify subtle expressions 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] 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.

[3] 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.

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

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

[6] 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)

[7] 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)

[8] P. Ekman, W.V. Friesen, “Unmasking the Face” (Malor Books).

[9] 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.