IEICE Proceeding Series

Facial Expression Recognition with Individual Adjustment

Koichi Takahashi, Yasue Mitsukura

Vol. 1 pp. 78-81
Publication Date: 2014/03/17
Online ISSN: 2188-5079

Facial Expression Recognition with Individual Adjustment

Koichi Takahashi† and Yasue Mitsukura†

†Keio University
3-14-1 Hiyoshi, Kohoku-ku, Yokohama, Kanagawa, Japan
Email: takahashi@mitsu.sd.keio.ac.jp

Abstract—In this paper, we propose a new method for individual adjustments of facial expression recognition. Facial expression recognition is not only a fundamental study in computer vision research, but also can be used to various applications such as avatar systems. We also aim to develop facial expression recognition methods for designing accurate and easy to use applications. However, individual adjustments are required for accurate recognition since each person has different shape of the face. Therefore, we propose a new method for individual adjustments of facial expression recognition. Our proposed method is able to calibrate individual differences easily by using an impersonal smiling intensity function and a scale factor which can impersonalize the personal properties.

1. Introduction

Facial expression recognition is a fundamental challenge in the research field of human-computer interaction and computer vision. This study is expected to be applied to the various applications such as facial expression mirroring for web conference tools.

Until now, many facial expression recognition methods are proposed. Accordingly, some effective applications based on facial expression recognition are recently available, for example, smiling recognition for digital cameras [1], emotional state analysis [2], and mirroring avatar system [3]. These studies achieved their purpose of applications by designing suitable technique for recognizing facial expressions.

In this paper, we aim to develop the new facial expression recognition method in order to design accurate and easy-to-use applications. The main target of the application is the avatar system for web conference tools. Some methods are previously proposed and commercially available. However, several problems still remain, such as improving accuracy and computing cost.

We have also proposed an avatar system using a monocular system [4]. Our method employs a cylindrical head model which has local templates placed at facial feature points. This method has merits as computing cost and accurate head pose estimation. In addition, by analyzing the deformations of facial feature points, facial expression and action can be automatically operated. However, this method requires the individual learning of radial basis function (RBF) network which is one of the neural networks [4]. This is due to the fact that each human has different shape of the face although our method utilizes the facial deformation as an input vector to the RBF networks.

Therefore, the objective of this study is to develop a new facial expression recognition method and individual adjustment technique. We especially address recognition of the smiling expression that is the most important expression for smooth human communications. In order to investigate the differences between smiling and speaking, the facial expression deformation measurement experiments are conducted. According to the results of 11 subjects, we show the importance of individual adjustment. The proposed method is able to easily calibrate individual differences using an impersonal smiling intensity function and a scale factor which can impersonalize the personal properties.

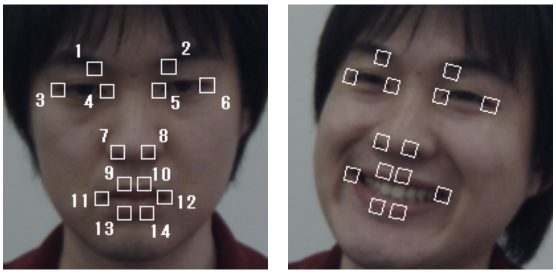
2. Head Pose Estimation

In our previous study, the simplified head model is proposed for the purpose of fast and accurate head pose estimation [5]. This head model is generated from a reference image as a user's head shape can be approximated by a cylinder. The simplified head model has 10 feature points: inner corners of both eyebrows, inner and outer corners of both eyes, both nostrils, and the outer corners of the mouth. In addition, reference templates are registered from the reference image and used for head pose tracking.

The simplified head model is a deformable model which originally has 3 degrees of freedom d_1, d_2, and d_3. Each parameter indicates the horizontal elastic movement of the mouth, the up-down movement of the mouth, and the up-down movement of the eyebrows, respectively. Therefore, the simplified head model is able to represent the user's facial expression deformation.

In this paper, we additionally set the feature points that are placed at 2 upper corners of the mouth and 2 lower corners of the mouth in order to extract lip shape features. Moreover, we add 2 degrees of freedom d_4 and d_5. Each parameter denotes the up-down movement of the upper and lower corner of the mouth, respectively. This simplified head model is shown in Fig. 1. The numbers in Fig. 1 (a) are the labels of feature points.

For the head pose estimation, we can apply particle filter [6] to the simplified head model. Particle filter is a time series signal processing method that can trace the target ob-



(a) Frontal view (b) Oblique view

Figure 1: Simplified head model.

jects according to the likelihood function. In this paper, we use the likelihood function based on template matching between the templates extracted from the reference image and the templates extracted from the frame image.

For the implementation of head pose tracking, we need to estimates the parameters of head position $[x\ y]^\top$, scale z, orientation $[yaw\ roll\ pitch]^\top$, and facial expression deformation $[d_1\ d_2\ d_3\ d_4\ d_5]^\top$. As a result, this becomes the 11 state tracking problem. The details of the simplified head model are mentioned in reference [5].

3. Facial Expression Measurement

By analyzing the facial expression deformation $[d_1\ d_2\ d_3\ d_4\ d_5]$, facial expression can be recognized. For example, smiling expression consists of the elastic movements of the mouth, and angry expression makes eyebrows furrowed. Therefore, detection of these facial feature changes and analysis of how these changes relate to typical facial expression are the main issues of this problem. However, existing individual differences of the head shape makes facial expression recognition difficult. This is because each human has different head shape, and thus, facial expression changes corresponding to typical facial expressions vary according to the persons. Therefore, accurate facial expression recognition requires the individual adjustments to calibrate these differences.

In order to investigate the individual differences of facial expression changes, we conducted the facial expression deformation measurement experiment. In this experiment, we especially address the recognition of smiling expression that is the most important expression for smooth human communications. Therefore, we measure the facial expression movements in the situations of "speaking" and "smiling".

This experiment is conducted by 9 male and 2 female subjects aged between 21 and 40. Each subject firstly sits down in front of a capture device. After the face is detected, subsequently, measurements of facial expression movements start. During the measurements, the subjects speak according to an alphabet song twice and face the indicated directions: "face to the left", "face to the right", "tilt to the left", "tilt to the right", "face forward", and

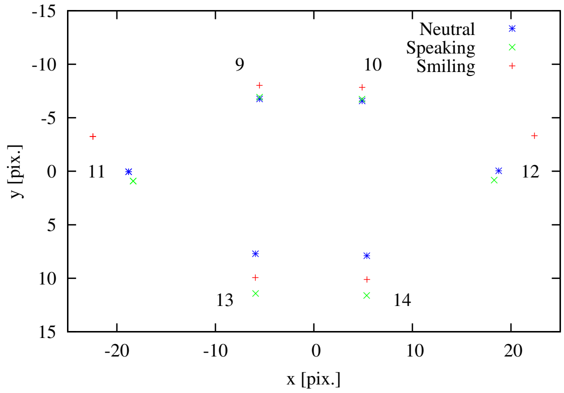


Figure 2: Average lip shapes.

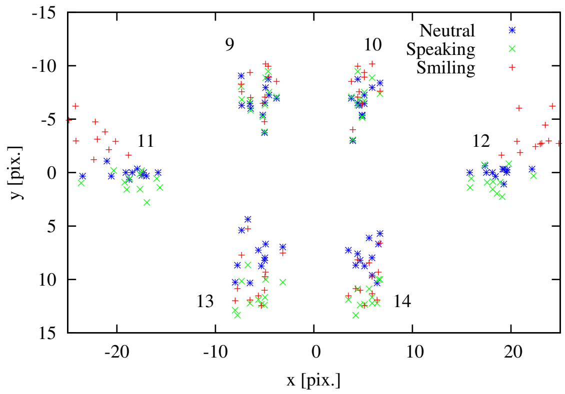


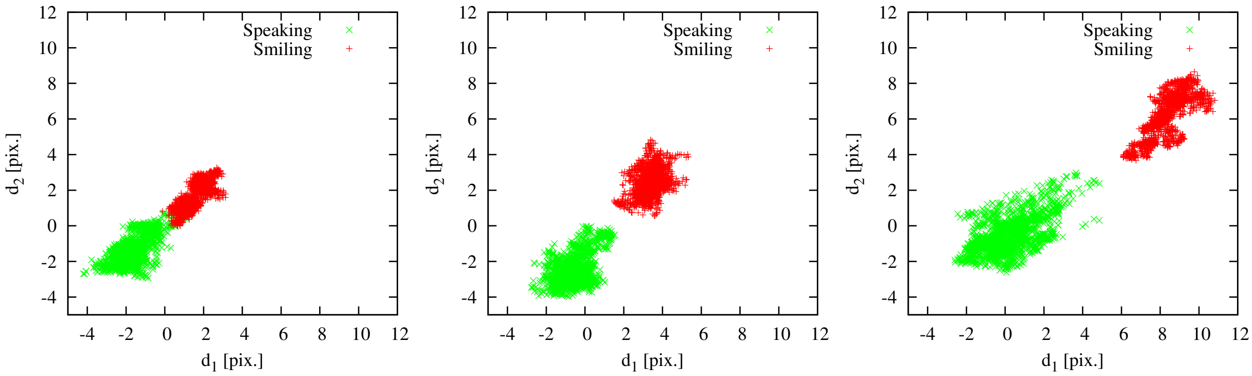
Figure 3: Individual differences of each lip shapes.

"face backward". Similarly, we measure the facial expression movements while each subject is smiling and facing above directions.

In this experiment, we utilize a PC with an Intel Core i7-2720QM 2.2GHz processor, Windows 7 OS, and 8GB of memory. We employ a Sony PlayStation Eye capture device that captures a 640×480 pixel resolution image at 60fps for the on-line video sequence. In this experimental setup, we can measure the person's facial expression deformations at 60 fps.

The results of average lip shapes are shown in Fig. 2, and the individual differences of the 11 subjects' lip shapes are depicted in Fig. 3. In these figures, the blue, green, and red results correspond to the neutral, speaking, and smiling expression, respectively. Each number in these figures corresponds to the labels in Fig. 1 (a). In addition, the scales of both axes are normalized by defining the distance between both eyes as 50[pix.].

As shown in Fig. 2, the average smiling lip shape varies from the neutral lip shape, especially both mouth corners. The average speaking lip shape also makes differences, especially toward the opening mouth. Above results indicate that the deformation of both mouth corners is the important for recognizing smiling expression. However, we can confirm that the lip shape has large individual differences as evidenced in Fig. 3. Each lip shape is widely



(a) Subject A (b) Subject B (c) Subject C

Figure 4: Individual differences of lip shape features among subjects.

distributed. These results show the difficulties of facial expression recognition without individual adjustments.

4. Individual Adjustment Method

The individual results of differences between speaking and smiling are illustrated in Fig. 4. In Fig. 4, we show the facial expression deformation results of the horizontal elastic movement of the mouth corner d_1 and the up-down movement of the mouth corner d_2 according to 3 subjects. The green points are the observed samples when the subject was speaking, and the red points are the observed samples when the subject was smiling. From left to right, the observed samples corresponding to speaking and smiling are distributed in wider range. This indicates that each person has different mouth shape properties, and thus, the calibration of each distribution is quite important. Therefore, we propose the new method for individual adjustment of facial expression changes.

We assume that the effective feature for recognizing smiling is based on the axis from the mean value of speaking distribution $\boldsymbol{\mu}_{speaking}$ to the mean value of smiling distribution $\boldsymbol{\mu}_{smiling}$ in the lip shape feature space. Here, lip shape features are referred to as $\boldsymbol{d} = [d_1\ d_2\ d_4\ d_5]^\top$. We can calculate a feature vector $\boldsymbol{v}$ by the following equation.

$$\boldsymbol{v} = \frac{\boldsymbol{\mu}_{smiling} - \boldsymbol{\mu}_{speaking}}{\left\|\boldsymbol{\mu}_{smiling} - \boldsymbol{\mu}_{speaking}\right\|} \tag{1}$$

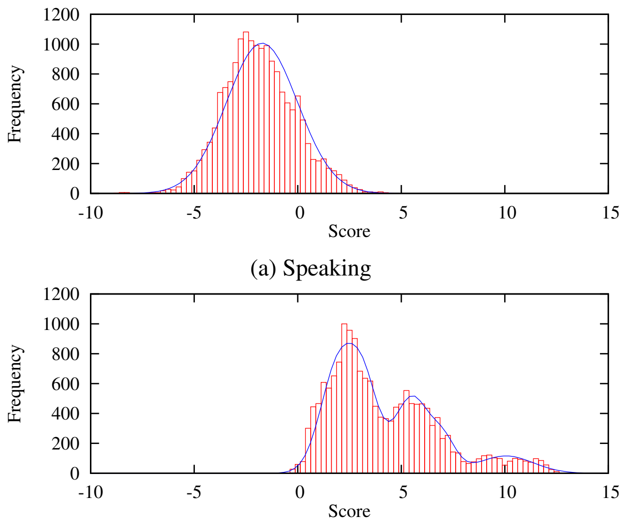
Here, $\|\cdot\|$ means a L^2 norm.

As a result, $\boldsymbol{v}$ is an unit vector of direction from the speaking distribution to the smiling distribution. By using $\boldsymbol{v}$, we can easily calculate the score $f(\boldsymbol{d})$.

$$f(\boldsymbol{d}) = \boldsymbol{v}^\top \boldsymbol{d} \tag{2}$$

This score means the smiling factor of the input vector $\boldsymbol{d}$.

The characteristics of this score $f(\boldsymbol{d})$ is shown in Fig. 5. In these figures, we depict the histograms of the speaking and smiling samples on the score function. Fig. 5 (a) is the result of speaking, and we can confirm the speaking samples are based on Gaussian distribution. In addition, we can consider this histogram is governed by the probability density function $P(f(\boldsymbol{d})|C = speaking)$, and $C \in \{speaking, smiling\}$ is a set of the class. Therefore, $P(f(\boldsymbol{d})|C = speaking)$ can be represented by the Gaussian function as illustrated by blue curve.



(a) Speaking

(b) Smiling

Figure 5: Histograms on the score function.

Similarly, the histogram of smiling is given in Fig. 5 (b). However, Fig. 5 (b) shows that $P(f(\boldsymbol{d})|C = smiling)$ is not based on the Gaussian distribution. This is perhaps due to the individual differences of smiling. In order to represent the $P(f(\boldsymbol{d})|C = smiling)$, we apply the Gaussian mixture model (GMM) to the observed samples of smiling. The result of GMM is illustrated by the blue curve in Fig. 5 (b). Here, the number of Gaussian is empirically set as 5.

By using above probability density functions, we can easily classify the input vector to speaking or smiling as

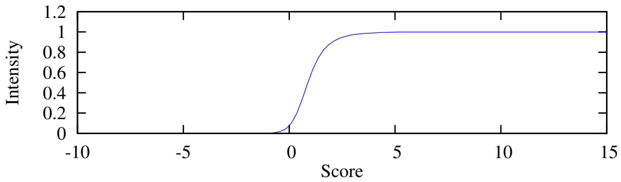


Figure 6: Impersonal smiling intensity $g(\boldsymbol{d})$.

Table 1: Recognition Results.

Input	Speaking[%]	Smiling[%]	Total[%]
Conventional	92.5	97.7	95.1
Proposed	93.9	99.5	96.6

following expression.

$$g(\boldsymbol{d}) = \frac{P(f(\boldsymbol{d})|C = smiling)}{\sum_{C\in\{speaking,smiling\}} P(f(\boldsymbol{d})|C)} \quad (3)$$

If this function outputs over 0.5, we classify the input $\boldsymbol{d}$ to smiling, if not, classify to speaking. This approach is equivalent to the maximum likelihood estimation of the facial expression recognition problem.

Incidentally, the property of $g(\boldsymbol{d})$ is shown in Fig. 6. As evidenced in this graph, the output starts to gradually increase around score equals to 0, and converges to 1. We define this function as an impersonal smiling intensity. Therefore, there is a need for an individual adjustment. In Fig. 4, we consider that the individual differences are governed by a scale factor in a direction toward vector $\boldsymbol{v}$. Hence, we use $g(\alpha\boldsymbol{d})$ instead of $g(\boldsymbol{d})$, and α is a scale factor for individual adjustment.

In this paper, we calculate α based on the mean of smiling distribution. Here, the impersonal smiling intensity $g(\boldsymbol{d})$ is designed where the mean of smiling distribution equals to $\boldsymbol{\mu}_{smiling}$. According to Fig. 4, the mean of personal smiling distribution $\boldsymbol{\mu}'_{smiling}$ is different from $\boldsymbol{\mu}_{smiling}$. The personal smiling distribution means the observed samples of each person such as red points in Fig. 4. We therefore set α as follows:

$$\alpha = \frac{f(\boldsymbol{\mu}_{smiling})}{f(\boldsymbol{\mu}'_{smiling})}. \quad (4)$$

As a result, α is able to conduct the individual adjustment by impersonalizing the personal smiling distribution.

For the comparison between the recognition results of conventional method and the proposed method, the recognition rate is shown in Table 1. In this paper, the recognition rate can be obtained by the number of correct samples divided by the number of target samples. The conventional method is referred to as the method without individual adjustments.

Table 1 shows that the proposed individual adjustment method can improve the accuracy in all cases. This is because the proposed method can appropriately adjust the decision boundary with the individual adjustments. The remaining errors are perhaps due to the problem of facial features which is used for recognition. For example, in Fig. 4 (a), there is overlap in the distributions of speaking and smiling. However, we can achieve 96.6% although only 4 features are used to represent the lip shape. We believe the proposed method is adequately practical for designing various applications based on facial expression recognition.

5. Conclusions

In this paper, we propose a new method for individual adjustments of facial expression recognition. Conventional facial expression recognition method can be used to various applications. However, individual adjustments are required for accurate recognition since each person has different appearance of the face. Therefore, we aim to propose a new method for individual adjustments of facial expression recognition.

Our proposed method is able to easily calibrate individual differences using an impersonal smiling intensity function and a scale factor which can impersonalize the personal properties. For the future works, we will modify our method to recognize various facial expressions and apply it to the applications such as mirroring avatar systems.

References

[1] J. Whitehill, G. Littlewort, I. Fasel, M. Bartlett, and J. Movellan, "Towards practical smile detection", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 31, No. 11, pp. 2106–2111, 2009.

[2] S. Kumano, K. Otsuka, D. Mikami, and Y. Junji, "Recognizing communicative facial expressions for discovering interpersonal emotions in group meetings", *International Conference on Multimodal Interfaces*, pp. 99–106, 2009.

[3] J. M. Saragih, S. Lucey, and J. F. Cohn, "Real-time avatar animation from a single image", *IEEE International Conference on Automatic Face and Gesture Recognition and Workshops*, pp. 117–124, 2011.

[4] K. Takahashi, M. Nakanishi, and Y. Mitsukura, "Recognizing facial actions using RBF networks", *2011 International Symposium on Nonlinear Theory and its Applications*, pp. 298–301, 2011.

[5] K. Takahashi, H. Fukai and Y. Mitsukura, "Facial expression estimation using simplified head model based on particle filtering," *The 11th IEEE International Workshop on Advanced Motion Control*, pp.173–178, 2010.

[6] M. Isard and A. Blake, "Condensation - conditional density propagation for visual tracking", *International Journal of Computer Vision*, Vol. 29, No. 1, pp. 5–28, 1998.