# Dynamics Analysis of Facial Expressions for Person Identification

Hidenori Tanaka and Hideo Saito

Graduate School of Science and Technology, Keio University
3-14-1, Hiyoshi, Kouhoku-ku, Yokohama, Kanagawa, 223-8522 Japan
hidenori@hvrl.ics.keio.ac.jp

**Abstract.** We propose a new method for analyzing the dynamics of facial expressions to identify persons using Active Appearance Models and accurate facial feature point tracking. Several methods have been proposed to identify persons using facial images. In most methods, variations in facial expressions are one trouble factor. However, the dynamics of facial expressions are one measure of personal characteristics. In the proposed method, facial feature points are automatically extracted using Active Appearance Models in the first frame of each video. They are then tracked using the Lucas-Kanade based feature point tracking method. Next, a temporal interval is extracted from the beginning time to the ending time of facial expression changes. Finally, a feature vector is obtained. In the identification phase, an input feature vector is classified by calculating the distance between the input vector and the training vectors using dynamic programming matching. We show the effectiveness of the proposed method using smile videos from the MMI Facial Expression Database.

**Keywords:** facial expression analysis, AAMs, LK-based feature point tracking, DP matching, person identification.

## 1 Introduction

Facial expression analysis is utilized in man-machine interfaces such as human-robot interactions. Most previous research in this field has tried to classify facial expressions into fundamental categories based on emotions. However, facial expressions contain not only expressions of emotions but also individual differences over time [1]. In this paper, we focus on the individual differences and propose a new method for analyzing the dynamics of facial expressions to identify persons.

To achieve biometric identification services, we consider that various physical features have to be fused and that facial expression is one of them. Since facial expression might not provide enough discriminating power, this research is considered as a type of soft biometrics. Figure 1 shows an example of using facial expressions for person identification service at a high-class membership club. At the entrance, a robot approaches the members and communicates with them while the identification process is performed.

In most person identification methods using facial images, the variations in facial expressions are one of the factors that lower the discriminating power.

X. Jiang and N. Petkov (Eds.): CAIP 2009, LNCS 5702, pp. 107–115, 2009.

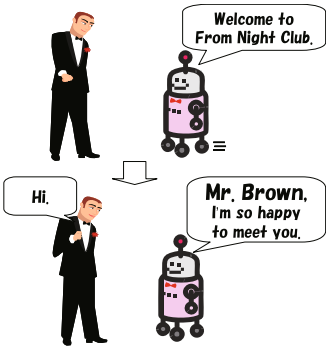



**Fig. 1.** Person identification service at a high class membership club

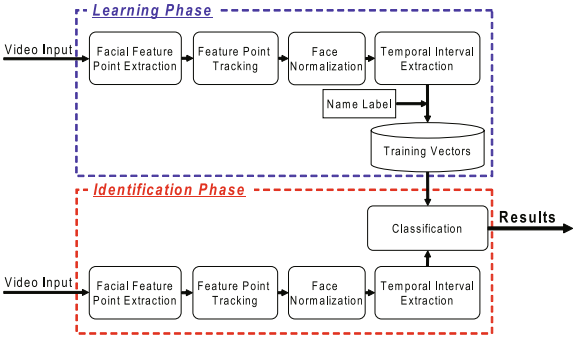



**Fig. 2.** Flow of proposed method

A number of methods [2, 3, 4] have been developed to address this problem. In the case of facial expression videos, however, person identification can be performed by using individual differences behind the facial expressions.

Previous research in the area of person identification using facial expressions is very limited. Ning et al. [5] generated features by summing up the flow fields over time using smile videos. In this method, the dynamics of facial expressions were not well described because features were generated by summing up the flow fields over time. Also, optical flow was calculated for whole facial images and did not accurately show the background regions of the images. This lowers the discriminating power. Further, facial feature points were manually extracted to normalize the facial images. Chen et al. [6] constructed a high-dimensional feature vector that concatenated a sequence of motion flow fields using videos of persons speaking. In this method, the dynamics of facial expressions were described. However, optical flow was calculated for whole facial images and facial feature points were manually extracted.

In contrast, we propose a novel method that analyzes the dynamics of facial expressions to identify persons using Active Appearance Models (AAMs) and accurate facial feature point tracking. The next section provides an overview of the proposed method and details its facial feature point extraction, feature point tracking, facial normalization, temporal interval extraction and identification processes. Section 3 demonstrates the method's effectiveness using smile videos from a published facial expression database. In Section 4, we offer conclusions pertaining to our work.

## 2 Proposed Method

The proposed method consists of two phases: "the learning phase," and "the identification phase." Figure 2 shows the flow of the method. In the learning

phase, facial feature points (eyebrow, eye, nose, mouth, and facial contour parts) are automatically extracted using AAMs in the first frame of the facial expression videos. The feature points are then tracked using the Lucas-Kanade based feature point tracking method (LK-based feature point tracking). Each facial image is normalized using three facial feature points to account for the variations in the object's head pose and those in the distance from face to camera. Temporal intervals are extracted from the difference between the feature points' position in the current frame and that in the previous frame. A feature vector is also generated and stored with the name label. In the identification phase, an input feature vector is generated as in the learning phase and classified by calculating the distance between the input vector and the training vectors using dynamic programming (DP) matching. The next subsections detail each process.

### 2.1 Facial Feature Point Extraction

To extract facial feature points in the first frame of each video, we use AAMs [7]. AAMs are generative and parametric models of a certain visual phenomenon that shows both shape and grey-level appearance variations. These variations are represented by a linear model.

Initially, grey-level variance independent from shape variance is needed for learning the correlation between shape and grey-level. The training data for AAMs is a set of images and coordinate values of feature points on the images. A shape vector $s$ is composed of coordinate values on feature points. A grey-level vector $g$ is composed of intensity values in a warped image, which is obtained by extracting the face region from an image along its feature points and normalizing its shape into a mean shape $\bar{s}$ of the normalized shapes.

Next, the distribution and correlation between shapes and grey-level is calculated. Principal Component Analysis (PCA) is performed on a set of shape vectors $s$ and grey-level vectors $g$ in training data.

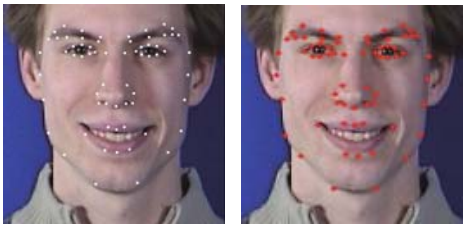
$$s = \bar{s} + P_s c_s, \; g = \bar{g} + P_g c_g \tag{1}$$

where $\bar{s}$ is a mean vector of $s$, $\bar{g}$ is a mean vector of $g$, $P_s$ and $P_g$ are orthogonal matrixes where each column vector is a base vector, and $c_s$ and $c_g$ are coefficients of the base vector. Since there may be correlations between the shape and grey-level variations, PCA is performed again. If an input image and the training model are given, we can treat facial feature point extraction as an optimization problem in which we minimize the grey-level difference between an input image and a synthesized image using the parameter vector $d^*$.

$$d^* = \arg\min_d \mid g_i - g_m \mid^2 \tag{2}$$

where $d$ is a parameter vector controlling both the shape and greylevels of the model, $g_i$ is a warped input image and $g_m$ is a synthesized image. $\mid g_i - g_m \mid$ are iteratively minimized and we get the optimization result $d^*$. From the vector $d^*$, the shape vector of an input image is obtained. Figure 3 shows an example of a training image. We put 65 feature points on the image.

**Fig. 3.** An example of a training image for AAMs



(a) AAM tracking (b) LK-based tracking

**Fig. 4.** Comparison of facial feature point tracking results

### 2.2 Feature Point Tracking

In each video, we need to track the facial feature points that are extracted in the first frame using AAMs. AAM tracking (part of AAM-API) is a training-based feature point tracking method in which it is necessary to make many training images to track the movement of facial feature points accurately. It is also known that AAM tracking based on a large number of datasets has difficulties in tracking accuracy because of local minimums [8]. In case we make training images consisting of a neutral face, the feature point tracking fails (around the mouth) as shown in Fig.4 (a). In our method, we use the feature point tracking method developed by Lien [9], which we call "LK-based feature point tracking," to track the feature points accurately. The goal of feature point tracking is to find the best matching positions between an $N$ x $N$ window $R$ in the $t$ frame and those in the $t+1$ frame that minimize the cost function $E$ of the weighted sum of squared differences (SSD) as follows:

$$E(\epsilon_x) = \sum_{x \in R} \omega_x \cdot [I_t(x) - I_{t+1}(x - \epsilon_x)]^2 \tag{3}$$

where $I_t(x)$ denotes the grey value of the pixel position $x$ in the $t$ frame, $\epsilon_x$ is the motion vector of $x$ between two consecutive frames, and $\omega_x$ is a window function for weighting the squared differences in $E$, which are defined by LK-based weight. Here, this weight is empirically determined. In our method, the feature point tracking is robustly performed against illumination variations because we use a part of the facial edge (facial feature points). The feature point tracking result using the LK-based feature point tracking method is shown in Fig.4 (b) . From this figure, we can see that the feature points are accurately tracked when the facial expression is changed.

### 2.3 Facial Normalization

To account for the object's head pose movements and the different distances from the object's face to the camera, each facial image is normalized. In our method, we use three facial feature points (two inner canthi and a philtrum)

to align each facial image. These three points are extracted in the first frame of each video using AAMs and then tracked using the LK-based feature point tracking method. These three points are then moved into the aligned positions. Facial images are aligned by 2D affine transformation with respect to these three points. The inner canthi and philtrum of all aligned facial images have the same coordinate values. After alignment, facial images are cropped and resized. In our experiment, the resized image size is 128 x 128 pixels.

### 2.4 Temporal Interval Extraction

To determine the frames that encompass the duration of facial expression changes, the starting and ending frames are extracted in each video. Frames that do not contain any facial expression changes will be abandoned because these frames will lower the discriminating power. In our method, the starting and ending frames are extracted by differences in the facial feature points between two successive frames as follows:

$$F(x^t) = \mid f(x^t) - f(x^{t-1}) \mid, \; f(x^t) = \sum_{i=1}^{K} \mid x_i^t - x_i^0 \mid \tag{4}$$

$$\begin{pmatrix} t_s = t & \text{if } F(x^t) > Th_{period} \\ t_e = t & \text{if } t > t_s \; and \; F(x^t) < Th_{period} \end{pmatrix}$$

where $F(x^t)$ denotes the differences between the coordinate values of the facial feature points in the $t$ frame and those in the $t-1$ frame, $K$ is the total number of facial feature points, $x_i^t$ denotes the coordinate values of the $i$th feature points in the $t$ frame, $t_s$ and $t_e$ are the starting and ending frames, and $Th_{period}$ is the threshold value. Here, $Th_{period}$ is empirically determined depending on the experimental environment because $Th_{period}$ is mainly affected by the illumination variations in the experimental environment. In our experiment, facial expression changes always start from a neutral face.

After this process, we can obtain a feature vector. It consists of the 2D coordinate sequence of facial feature points and the vector dimensions total 2D coordinate x 65 points x ($t_e$-$t_s$) frames.

### 2.5 Identification

Identification is performed by classifying an input feature vector. Because temporal intervals of facial expression changes will not be the same between different individuals and will not be constant at all times even for the same person, we need to absorb the variations. There are many matching algorithms to compare the patterns whose temporal intervals are different. In our method, we use conventional dynamic programming (DP) matching to compare an input feature vector with the training feature vectors. In detail, the distance $G(i, j)$ between an input feature vector $A = (a_1, a_2, \ldots, a_i, \ldots, a_{T_1})$ and the training feature vector $B = (b_1, b_2, \ldots, b_j, \ldots, b_{T_2})$ is calculated as follows:

$$G(i,j) = \min \begin{pmatrix} G(i-1,j) + D(i,j) \\ G(i-1,j-1) + 2D(i,j) \\ G(i,j-1) + D(i,j) \end{pmatrix} \tag{5}$$

where $D(i,j)$ denotes the Euclidian distance between $a_i$ and $b_j$. The calculated distance is normalized by the length of the input vector and the training vector $(T_1 + T_2)$ and the DP distance is obtained. In identification, the input vector is classified by the threshold value.

## 3 Experiments

To show the effectiveness of our method, we conducted experiments using smile videos from the MMI Facial Expression Database [10]. In our experiments, the resolution of the videos is 720 x 576 pixels and the frame rate is 25 frames/second. Facial expression changes start with a neutral face, move to a smile, and then go back again to the neutral face. We selected 48 smile videos (12 persons, 4 videos/person) from the database to evaluate. All videos were very nearly frontal facial images, and so ideally suit our facial normalization method.

We first evaluate the discriminating power of our method using all facial feature points, comparing them with the previous method. Then, we evaluate the discriminating power of our method for each facial part. For evaluation purposes, we considered that one video was for test data and that the other videos were for training data. To evaluate the discriminating power, we used the equal error rate (EER) and the recognition rate (RR). EER is the probability that the false acceptance rate (FAR) and the false rejection rate (FRR) are equal. In general, the discriminating power is high when EER is lower and RR is higher.

### 3.1 Discriminating Power of a Whole Face

In this experiment, we first show the discriminating power of our method with all facial feature points (65 points) using smile videos. Figure 5 shows the tracking results of three persons in smile videos. From this figure, we can see individual differences in smile dynamics. From the evaluation results, the EER value was 14.0% and the RR value was 92.5%. This shows that smile dynamics represented by our method have high discriminating power.

To compare the discriminating power of our method with that of the previous method, we applied the optical-flow based method to the same datasets. In this experiment, each facial image was sampled into 64 points evenly spaced over a whole facial image for computing the optical flow field, and optical-flow was calculated against the points. A feature vector was obtained after temporal interval extraction. The identification was performed in the same way as in our method. From the evaluation results, the EER value was 36.8% and the RR value was 62.3%. These results show that the discriminating power of our method is higher than that of the optical-flow based method. Here, the factor that lowers the discriminating power in the optical-flow based method was some

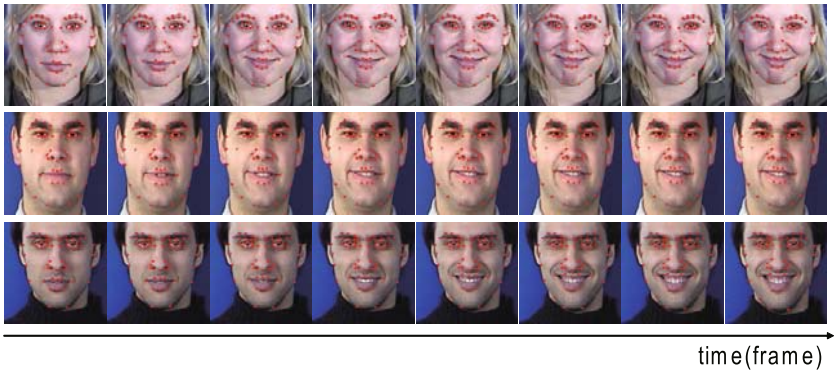



**Fig. 5.** Tracking results of three persons in smile videos (Every three frames from a neutral face to a smile face)

unexpected flows found in the background regions of the facial image. In contrast, in our method, we extract the facial feature points of the facial image and the background regions of the image do not affect the tracking process.

### 3.2 Discriminating Power in Each Facial Part

In this experiment, we show the discriminating power of our method in each facial part (eyebrow, eye, nose, mouth, and facial contour parts) using smile videos. We also show the differences in the discriminating power by temporal interval extraction because temporal interval plays an important role in the dynamics of facial expressions. First, we use the same temporal interval as in the previous experiments (temporal interval extraction by all facial feature points). Table 1 shows EER and RR results obtained for each facial part. From this table, we can see that the EER value of the eyebrow part is lowest and the RR value of the eyebrow part is highest. From this result, it can be said that the eyebrow part had higher discriminating power than the other parts in these datasets. On the other hand, we can also see that the EER values of the eye part and those of the mouth part are higher and that the RR values of the eye part and those of the mouth part are lower. From this result, it can be said that the eye and mouth parts have less discriminating power than the other parts, while wrinkling around the corners of the eyes and a rising in the corners of the mouth are characteristic movements in smiles.

Next, we use the temporal interval extracted by each facial feature point. Table 2 shows EER and RR results obtained for each facial part. From this table, we can see that the EER values of all parts in Table 2 are higher than those in Table 1 and the RR values of all parts in Table 2 are mostly lower than those in Table 1. From this result, it can be said that temporal interval extraction by all facial feature points had higher discriminating power than that by each facial

**Table 1.** EER and RR in each facial part (Temporal interval is extracted by all facial feature points)

| facial part | EER[%] | RR[%] |
|---|---|---|
| eyebrow | 15.2 | 88.7 |
| eye | 24.4 | 71.7 |
| nose | 21.7 | 70.7 |
| mouth | 24.7 | 64.2 |
| facial contour | 17.0 | 81.1 |

**Table 2.** EER and RR in each facial part (Temporal interval is extracted by each facial feature points)

| facial part | EER[%] | RR[%] |
|---|---|---|
| eyebrow | 17.3 | 77.4 |
| eye | 31.0 | 67.9 |
| nose | 22.5 | 81.1 |
| mouth | 27.6 | 62.3 |
| facial contour | 19.0 | 67.0 |

feature point. In other words, the combination of facial parts' movements over time had individual differences.

## 4 Conclusions

In this paper, we proposed a method for analyzing the dynamics of facial expressions to identify persons. We automatically extracted facial feature points and accurately tracked them. We evaluated the discriminating power of our method using 48 smile videos from the MMI Facial Expression Database. The evaluation results showed that the EER value was 14.0% and the RR value was 92.5% and the discriminating power of our method was higher than that of the previous method. We also found that the eyebrow part had higher discriminating power than the other parts of the face and that the eye and mouth parts had less discriminating power than the other parts even though these parts are characteristic parts in smiles. Further, the combination of facial parts' movements over time had individual differences.

In future work, we plan to generate a feature vector while considering the appearance of facial images and evaluate our method using other facial expression videos.

## References

1. Cohn, J.F., Schmidt, K., Gross, R., Ekman, P.: Individual differences in facial expression: stability over time, relation to self-reported emotion, and ability to inform person identification. In: ICMI 2002, pp. 491–496 (2002)
2. Tsai, P.H., Jan, T.: Expression-invariant face recognition system using subspace model analysis. In: SMC 2005, vol. 2, pp. 1712–1717 (2005)
3. Ramachandran, M., Zhou, S.K., Jhalani, D., Chellappa, R.: A method for converting a smiling face to a neutral face with applications to face recognition. In: ICASSP 2005, vol. 2, pp. 977–980 (2005)
4. Li, X., Mori, G., Zhang, H.: Expression-invariant face recognition with expression classification. In: CRV 2006, pp. 77–83 (2006)
5. Ning, Y., Sim, T.: Smile, you're on identity camera. In: ICPR 2008, pp. 1–4 (2008)

6. Chen, L.-F., Liao, H.-Y.M., Lin, J.-C.: Person identification using facial motion. In: ICIP 2001, vol. 2, pp. 677–680 (2001)
7. Cootes, T.F., Edwards, G.J., Taylor, C.J.: Active appearance models. IEEE Transactions on PAMI 23(6), 681–685 (2001)
8. Gross, R., Matthews, I., Baker, S.: Generic vs. person specific active appearance models. Image and Vision Computing 23, 1080–1093 (2005)
9. Lien, J.: Automatic recognition of facial expressions using hidden markov models and estimation of expression intensity. PhD thesis, Carnegie Mellon University (1998)
10. Pantic, M., Valstar, M., Rademaker, R., Maat, L.: Web-based database for facial expression analysis. In: ICME 2005, pp. 317–321 (2005)