Proceedings of the 2009 IEEE International Conference on Systems, Man, and Cybernetics
San Antonio, TX, USA - October 2009

Expression-invariant Facial Identification

Pohsiang Tsai
School of Computing and Communications
Faculty of Engineering and Information Technology
University of Technology, Sydney, Australia
pohsiang@it.uts.edu.au

Tich Phuoc Tran
School of Computing and Communications
Faculty of Engineering and Information Technology
University of Technology, Sydney, Australia
tiptran@it.uts.edu.au

Longbing Cao
School of Software
Faculty of Engineering and Information Technology
University of Technology, Sydney, Australia
lbcao@it.uts.edu.au

***Abstract*—Facial identification has been recognized as most simple and non-intrusive technology that can be applied in many places. However, there are still many unsolved facial identification problems due to different intra-personal variations. In particular, when images of the databases appear at different facial expressions, most currently available facial recognition approaches encounter the expression-invariant problem in which neutral faces are difficult to be recognized. In this paper, a new approach is proposed to transform facial expressions to neutral-face like images; hence enabling image retrieval systems to robustly identify a person's face for which its learning and testing face images differ in facial expression.**

***Keywords*—Face Recognition, Expression-Invariant Face Recognition, Pseudo-Neutral Image**

I. INTRODUCTION

Facial recognition has received substantial attention in the biometrics. Although a great deal of effort has been devoted to automated facial recognition systems, it still remains a challenging problem. This is because human facial appearance has potentially very large intra-subject variations due to head pose, illumination, facial expression, occlusion due to other objects or accessories, facial hair and aging. According to Adini et al. [1], it is desirable to have a recognition system which is able to recognize a face insensitive to these within-personal variations. An expression-invariant solution is needed when a facial recognition system only has neutral images for training and expressive images for testing. That is, the recognition performance trained only with neutral facial expressions will drop if there are facial expression changes in the appearance of facial images in probe images. In this paper, we presented a pre-processing approach to transform facial expressions to pseudo neutral facial images. This enables facial image retrieval systems to robustly identify a person's face for which its learning and testing face images differ by facial expression. We found that our conversion method brought facial expressions close to their corresponding neutral faces; hence, enabling the capability of a learning framework to discern among human faces.

II. RELATED WORK IN EXPRESSION-INVARIANT FACIAL RECOGNITION

When we want to identify a person's identity, faces with facial expressions make the recognition more difficult [2]. Facial expression invariant recognition is still an unsolved problem. Facial expression images can be morphed into pseudo images that are similar to neutral images and are the same as the ones used for training. However, there is no guarantee that all images can be morphed correctly due to lack of facial texture information [3]. For example, lack of texture inside closed eyes may affect the morphing from facial expression images into neutral images. An alternative is to learn the local motions within facial images that represent facial expression changes of faces. However, individuals express their own emotions or expressions in somewhat different ways [3] (especially different people with various cultural backgrounds). That means this technique will falter because we cannot morph correctly. Of course, there are other ways to cope with expression variant problems. Examples of research in expression-invariant facial recognition are as follows:

In [4], a quantified statistical facial asymmetry method under 2D facial expression changes (called AsymFaces) was used for person identification. PCA was then applied to AsymFaces for dimension reduction. AsymFaces was claimed to be invariant to facial expression changes. In [5], a local and probabilistic weighting method that weights the local areas of facial features independently was used. This method is less sensitive to expression changes. In [6], the authors extended a previous local probabilistic approach presented by Martinez [5], using Self-Organizing Map (SOM) instead of a mixture of Gaussians to learn the facial subspace that represented each individual. They used a soft k Nearest Neighbor (soft k-NN) ensemble method as a decision maker. Their intention was to tackle the 'one sample problem' and the expression-invariant problem. They used 'neutral' for training and testing separately with 'smile', 'anger', and 'scream' faces. They claimed that their proposed method gave high robust performance against the variant expressions. However, contrary to their claim the performance was degraded to an average of 14.7% in the second session (two weeks later). In [3], an Adaptive Principle

Component Analysis (APCA) method was used to deal with one sample problem under both illumination and facial expression changes simultaneously. The APCA method was applied to 2D face images after applying standard PCA method in order to construct a subspace for image representation and to improve class separability. A Bayes classifier was then used for classification. In [7], the authors used preprocessing steps for converting a smiling face to a neutral face by using a triangular mesh model (CANDIDE) for generic human face and Active Appearance Models (AAMs) to find the optimal shape and appearance parameters for best describing a face. Once these two geometry meshes have been registered, they changed the geometry mesh of expressions to their neural faces. The texture information of the new changed mesh was then mapped from the original mesh using an affine warping of the texture. They then used a PCA+LDA classifier for recognition. Their constraints were that only expressions concerning different configurations of the mouth were used. It might be computationally expensive to generate the mesh models.

However, their appearance-based approaches still suffer from the high dimensionality problem; that is to say, this problem will require expensive computation and increase the sparse data distribution. The abovementioned techniques henceforth were developed to solve appearance-based facial recognition problems under facial expression changes; these changes would affect the pixel-based values, and these values did not convey the dynamic changing information of facial expression changes. In this paper, we will use distance-based facial features of facial fiducial points instead.

III. Proposed Method

Recognition performance of facial recognition system that trained with only neural faces will drop if there are facial expression variations in the appearance of facial images. That is, when images of the databases appear at different facial expressions, most currently available facial recognition approaches encounter the expression-invariant problem in which neutral faces are difficult to recognize.

In facial recognition, we only consider facial expressions as the abnormal behaviors in comparison with expressionless faces as the normal behaviors. Our aim is to propose a boosting framework that first converts different expression changes (abnormal behavior) into their corresponding pseudo neutral facial images in the preprocessing stage and then adopts the AdaBoost.M1 algorithm [8] [9] to do the classification. The conversion method in the preprocessing stage is motivated by the idea of image segmentation [10] [11]. In image segmentation the aim is to find region of interests (ROIs) in any image; that is, to segment any ROI (foreground) from an image common background. Accordingly, we will treat neutral faces as mean faces (background) and facial expression as variance image (foreground) that deviates from its mean face (neutral face). Therefore, our aim is to restore one's neutral face when one displays his/her deviated face (facial expression).

Suppose that we have input image data set X, which consists of neu_i^k and ex_j^k; that is, $X = neu_i^k + \exp_j^k$ where $neu^k(i)$ stands for expressionless faces of a class k and $\exp_j^k$ stands for expression faces of a class k. This proposed framework of a preprocessing and classification stage is explained in the following sections.

A. Preprocessing Stage

In the preprocessing stage, we convert any facial expression image $\exp_j^k(x,y)$ in below into its corresponding individual neutral facial image $neu_i^k(x,y)$ so as to yield a compensated facial image $SimFace_j^k(\exp_j^k(x,y), neu_i^k(x,y))$ that looks similar to its original neutral facial image $neu_i^k(x,y)$ as follows:

$$\begin{aligned} & SimFace_j^k\left(\exp_j^k(x,y), neu_i^k(x,y)\right) \\ & = ComFace_j^k\left((x,y), neu_i^k(x,y)\right) \\ & = \Delta_j^k + \mu^k \\ & = \exp_j^k(x,y) - neu_1^k(x,y) + \mu^k \\ & = \exp_j^k(x,y) - \eta \end{aligned} \tag{1}$$

, where $\Delta_j^k = \exp_j^k(x,y) - neu_1^k(x,y)$ is the image intensity difference between any facial expression image of the kth individual and its corresponding base neutral facial image of the same kind; μ^k is the mean neutral facial image of the kth individual; and $\eta = (neu_1^k(x,y) - u^k)$ is the offset neutral image served as the base image.

Once any facial expression images have been converted into their corresponding pseudo neutral facial images, then we use the AdaBoost.M1 algorithm to do the recognition.

B. Classification Stage

AdaBoost.M1 is the first straightforward extension of the AdaBoost algorithm. Suppose that we have the set of all possible k finite labels $Y = \{1, 2, ..., k\}$ for multiple classification tasks, and a set of all possible t finite weak learners $T = \{1, ..., t\}$ for a final single composite (stronger) learner $h_f = X \rightarrow Y$ that outputs the final classification decisions. The training error is defined as $E_t = \sum_{h_t(x_i) \neq y_i} W_t / \sum_{i=1}^{N} W_i^t$. The weight is updated by $W_{t+1}(x_i) = W_t(x_i) \times \beta_t^{1 - [\![h_t(x_i) \neq y_i]\!]}$, where $\beta_t = E_t / (1 - E_t)$. The final composite classifier $h_f(x) = \arg \max_{\zeta_t(x) = \omega_k} \left(\sum_{t=1}^{T} \left(\log \frac{1}{\beta_t} \beta_t [\![h_t(x) = y]\!] \right) \right)$ yields the output decision $Y = \{1, 2, ..., k\}$. Figure 1 shows this algorithm.

IV. EXPERIMENTSL RESULTS

One of the databases of facial expression images used in this article is the Japanese Female Facial Expression (JAFFE) database[1]. This database is used for facial expression analysis and recognition [12] [13] [14]. It contains 213 gray-scale images of 6 facial expressions (happiness, sadness, surprise, anger, disgust, fear) plus one neutral face posed by 10 Japanese women. Each woman posed for two, three or four examples of each of the six basic facial expressions and a neutral face. Each individual pose is for various extents of facial expression

Input:

- Set of N training sequence of examples $\langle (x_1, y_1), \ldots, (x_N, y_N) \rangle \in \Re^{N \times m}$ with their corresponding labels $y_i \in Y = \{1, \ldots, k\}$ for multiple classes
- Distribution (weight) W over the N examples
- Weak learning algorithm **WeakLearner** h_t
- An integer T specifying number of iterations

Output:

- Calculate the support for class ω_k by $h_f(x) = \arg \max_{\zeta_t(x)=\omega_k} \left(\sum_{t=1}^{T} \left(\log \frac{1}{\beta_t} \right) \beta_t \left[h_t(x) = y \right] \right)$, then the class with the maximum support is chosen as the label for x.

Algorithm:

1. **Initialize** the parameters

- The weight vector $W_i^1(x_i) = \frac{1}{N} \in [0,1]$ for $i = 1, \ldots, N$ and $\sum_{i=1}^{N} W_i^1 = 1$
- Set the initial ensemble $H = \varnothing$
- Set T, the number of classifiers to train
- Set initial values for all errors $E_{i|i=1,\ldots,N}$ of $h_{t|1,\ldots,T}$

2. **Do while** ($t \leq T \wedge E_t < 0.5 \wedge E_t \neq 0$)

- Call and train the WeakLearner h_t, providing it with the sample distribution W_t using the resampling technique, to get back $h_t : X \to Y$ with the training error E_t of h_t as defined in $E_t = \sum_{h_t(x_i) \neq y_i} \frac{W_t}{\sum_{i=1}^{N} W_i^t}$.
- Update the individual weights W_t by $W_{t+1}(x_i) = W_t(x_i) \times \beta_t^{1-[h_t(x_i) \neq y_i]}$
- Set $t = t + 1$

Figure 1. Adaboost.M1 Algorithm

changes. Some individuals pose similar facial expression changes, while others pose different ones. Also, some individuals have slight pose variations when they pose facial expressions. The size of each image is 256x256, resulting in an input dimensionality of $d = 65536$. Figure 2 displays the sample images in the database.

In this article, we used the distance-based facial features instead of using pixel-based facial appearance features to represent the dynamic changing factors of facial expression changes. The distance-based facial features are chosen based on facial fiducial points. These points are selected similar to [15] . The distance-based facial features are selected as the descriptors of faces for recognition [16] [17]. We then manually selected 18 fiducial characteristic points on each of the images for representing the original 17 Euclidean distance-based facial features superimposed on the subject's face image in Figure 3. Note that we chose ‘a’ as the base point because that point does not move when changing expressions. These

[1] http://www.kasrl.org/jaffe.html

Figure 2. Facial Expression Images from JAFFE database.

features are marked as F1, F2 … F17 (See TABLE 1). Hence, these facial features provided certain discriminative information when individuals change expressions.

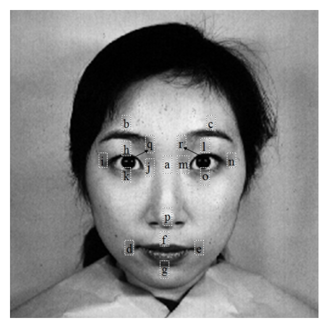

Figure 3. The 18 fiducial points on the subject’s face image.

TABLE 1. THE ORIGINAL 17 PRE-SELECTED FACIAL FEATURE DISTANCES.

Features	F1	F2	F3	F4	F5	F6	F7	F8	F9
Distances	a_b	a_c	a_d	a_e	a_f	a_g	a_h	a_i	a_j
Features	F10	F11	F12	F13	F14	F15	F16	F17	
Distances	a_k	a_l	a_m	a_n	a_o	a_p	a_q	a_r	

A. *Facial Expression-Invariant Recognition*

The goal of this section was to test how well our proposed framework performed on images from a database of frontal facial expression images. We applied this framework to ten subjects of the JAFFE database. The neutral facial images were used for training, while the others for identification.

Figure 4 shows the conversion results of the preprocessing method. This method converts all the different facial expressions into their corresponding pseudo facial images. These converted images are then used in the recognition stage.

The geometric-based (Euclidean distance-based facial fiducial points) facial features of facial neutral and pseudo neutral images will be used in the recognition stage.

B. Recognition Performance Comparision

In this section, we will show the experimental results of five different approaches. Figure 5 shows the recognition performances of different approaches. EigenExpression and FisherExpression showed the low recognition rates before converting to their pseudo facial images. Their performances were 33% and 40% (when number of features is 9), respectively. SimNeu using PCA and SimNeu using LDA showed increased recognition rates after using pseudo facial images as inputs. Their performances were 41% and 57% (when the number of features is 9), respectively. Although the SimNeu using PCA had slightly increased performance, the SimNeu using LDA had approximately 20% increased performance. The findings showed that their pseudo facial images had become closer to their corresponding neutral faces. However, due to the limitations of the standard algorithms, the recognition performances of pseudo facial images only showed slight improvement. Another reason could be because the human facial structures were not very different. The Euclidean distance-based approach using facial fiducial points could only carry precise individual differences to a limited extent. Therefore, an advanced method should be employed to be able to distinguish the differences among human faces.

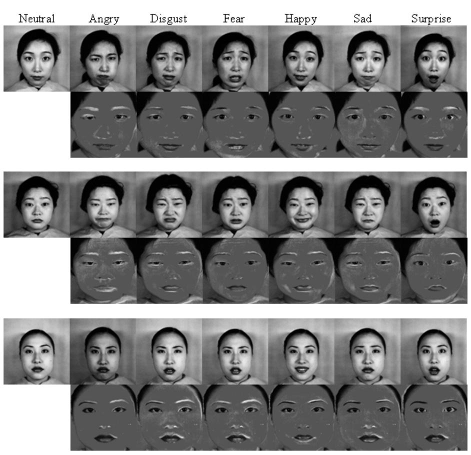

Figure 4. Pseudo facial images of different facial expressions after the preprocessing stage. The second row for each subject shows the pseudo facial images.

The proposed framework using AdaBoost algorithm not only uses the pseudo facial images, but is also capable of distinguishing even the minor distance differences among human faces. Figure 5 shows the proposed method had dramatically increased performance rate (96% when the number of features is 9). Clearly, it indicates that the facial expressions that been converted into pseudo facial images before the recognition can bring the facial expressions close to neutral faces. By using the AdaBoost framework, it can provide a boosting recognition capability of discerning among human faces.

V. CONCLUSION

This paper examines the issue of expression-invariant facial recognition by using an image conversion method to transform facial expressions to pseudo neutral images in the pre-processing stage and a boosting algorithm to do the classification. This conversion method was motivated by image segmentation in image processing.

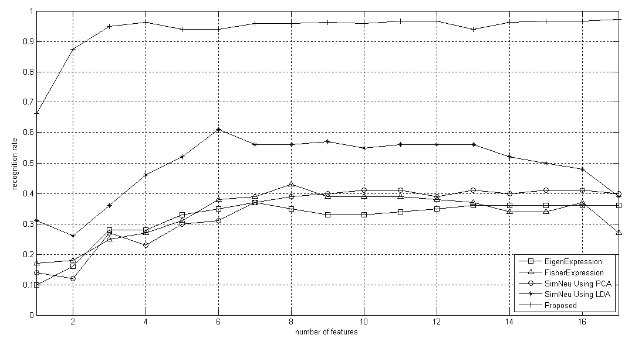

Figure 5. Recognition performances of different approaches.

We transformed the facial expression images into their corresponding pseudo neutral facial images. We then used the Euclidean distance-based facial features of 18 fiducial points on facial neutral and pseudo neutral images in the classification stage. The experimental results showed that the proposed framework provided a boosting recognition capability for discerning among human faces on the pseudo neutral images. The reasons that our proposed method outperformed other methods were 1) facial expressions in the testing set were converted into their corresponding pseudo neutral images before testing; making them closer to their corresponding neutral faces and 2) the AdaBoost.M1 algorithm was capable of dealing with hard-to-classify patterns.

The experimental showed that the proposed method had an increased performance rate. It is possible that the pseudo facial images of the converted facial expressions had become closer to the corresponding neutral faces.

REFERENCES

[1] Y. Adini, Y. Moses, and S. Ullman, "Face recognition: the problem of compensating for changes in illumination direction," *IEEE Transactions on Pattern Analysis and Machine Intelligence,* vol. 19, pp. 721-732, July 1997.

[2] A. M. Martinez, M.-H. Yang, and D. J. Kriegman, "Special issue on face recognition," *Computer Vision and Image Understanding: CVIU,* vol. 91, pp. 1-5, 2003.

[3] S. Chen and B. C. Lovell, "Illumination and expression invariant face recognition with one sample image," in *Proceedings of the 17th International Conference on Pattern Recognition*, 2004, pp. 300-303.

[4] Y. Liu, K. L. Schmidt, J. F. Cohn, and S. Mitra, "Facial asymmetry quantification for expression invariant

human identification," *Computer Vision and Image Understanding: CVIU,* vol. 91, pp. 138-159, 2003.

[5] A. M. Martinez, "Recognizing imprecisely localized, partially occluded, and expression variant faces from a single sample per class," *IEEE Transactions on Pattern Analysis and Machine Intelligence,* vol. 24, pp. 748-763, 2002.

[6] X. Tan, S. Chen, Z.-H. Zhou, and F. Zhang, "Recognizing partially occluded, expression variant faces from single training image per person with SOM and soft k-NN ensemble," *IEEE Transactions on Neural Networks,* vol. 16, pp. 875-886, 2005.

[7] M. Ramachandran, S. K. Zhou, D. Jhalani, and R. A. C. R. Chellappa, "A method for converting a smiling face to a neutral face with applications to face recognition," in *IEEE International Conference on Acoustics, Speech, and Signal Processing*, 2005, pp. 977-980 Vol. 2.

[8] Y. Freund and R. E. Schapir, "A decision-theoretic generalization of on-line learning and an application to boosting," *Journal of Computer and System Sciences,* vol. 1, pp. 119 - 139, 1997.

[9] L. I. Kuncheva, *Combining Pattern Classifiers: Methods and Algorithms*: John Wiley and Sons, 2004.

[10] M. Sonka, V. Hlavac, and R. Boyle, *Image Processing, Analysis, and Machine Vision*. New York: PWS

[11] S. E. Umbaugh, *Computer Imaging - Digital Image Analysis and Processing*. New York: CRC Press, 2005.

[12] M. Lyons, S. Akamatsu, M. Kamachi, and J. Gyoba, "Coding facial expressions with Gabor wavelets," in *IEEE International Conference on Automatic Face and Gesture Recognition*, Nara 1998, pp. 200-205.

[13] Z. Zhang, M. Lyons, M. Schuster, and S. Akamatsu, "Comparison between geometry-based and Gabor-wavelets-based facial expression recognition using multi-layer perceptron," in *IEEE International Conference on Automatic Face and Gesture Recognition*, Nara, 1998, pp. 454-459.

[14] M. J. Lyons, J. Budynek, and S. Akamatsu, "Automatic classification of single facial images," *IEEE Transactions on Pattern Analysis and Machine Intelligence,* vol. 21, pp. 1357-1362, Dec. 1999.

[15] M. Pantic and L. J. M. Rothkrantz, "Facial action recognition for facial expression analysis from static face images," *IEEE Transactions on Systems, Man and Cybernetics, Part B.,* vol. 34, pp. 1449-1461, June 2004.

[16] S. Z. Li and J. Lu, "Face Recognition Using the Nearest Feature Line Method," *IEEE trans. Neural Networks,* vol. 10, pp. 439-443, March 1999.

[17] M. Valstar and M. Pantic, "Fully Automatic Facial Action Unit Detection and Temporal Analysis," in *2006 Conference on Computer Vision and Pattern Recognition Workshop*, 2006, pp. 149-149.