



Queensland University of Technology
Brisbane Australia

This is the author's version of a work that was submitted/accepted for publication in the following source:

Wibowo, Moh Edi, Tjondronegoro, Dian W., & Zhang, Ligang (2012) On the application of the probabilistic linear discriminant analysis to face recognition across expression. In *2012 IEEE International Conference on Multimedia and Expo Workshops*, IEEE Computer Society, Melbourne, Australia, pp. 459-464.

This file was downloaded from: <http://eprints.qut.edu.au/49485/>

© Copyright 2012 IEEE Computer Society

2012 IEEE. Personal use of this material is permitted. Permission from IEEE must be obtained for all other uses, in any current or future media, including reprinting/republishing this material for advertising or promotional purposes, creating new collective works, for resale or redistribution to servers or lists, or reuse of any copyrighted component of this work in other works. Must include either a link to the abstract of the published article in IEEE Xplore, or the article's Digital Object Identifier (DOI). Reference: http://www.ieee.org/publications_standards/publications/rights/rights_policies.htm

Notice: *Changes introduced as a result of publishing processes such as copy-editing and formatting may not be reflected in this document. For a definitive version of this work, please refer to the published source:*

<http://dx.doi.org/10.1109/ICMEW.2012.86>

On the Application of the Probabilistic Linear Discriminant Analysis to Face Recognition across Expression

Moh Edi Wibowo, Dian Tjondronegoro, Ligang Zhang

Mobile Innovation Lab
Queensland University of Technology
Brisbane, Australia

moh.wibowo@student.qut.edu.au, dian@qut.edu.au, ligzhang@gmail.com

Abstract— Facial expression is one of the main issues of face recognition in uncontrolled environments. In this paper, we apply the probabilistic linear discriminant analysis (PLDA) method to recognize faces across expressions. Several PLDA approaches are tested and cross-evaluated on the Cohn-Kanade and JAFFE databases. With less samples per gallery subject, high recognition rates comparable to previous works have been achieved indicating the robustness of the approaches. Among the approaches, the mixture of PLDAs has demonstrated better performances. The experimental results also indicate that facial regions around the cheeks, eyes, and eyebrows are more discriminative than regions around the mouth, jaw, chin, and nose.

Keywords—face recognition; expression-invariant; probabilistic linear discriminant analysis

I. INTRODUCTION

With the success of face recognition methods on passport-photograph-like images [1–4], face recognition studies are now directed more to deal with uncontrolled environments. The current researches seek methods which are robust to variations of pose, illumination, expression, and age as well as to the presence of occlusion and low image resolution. Compared to recognition across pose and illuminations, recognition across expressions have been less explored and investigated.

Facial expressions are caused by movements of facial muscles and convey the emotional states of an individual. To formally describe these expressions, a model called Facial Animation Parameter has been developed under the MPEG-4 Facial Animation Coding System. Facial expressions can significantly affect the performances of face recognition algorithms [5, 6] particularly when the probe and gallery images have different expressions. Expression differences can be regarded as deformations between faces. The bigger the deformations, the harder the recognition will be. That’s why “extreme” expressions such as screaming are of the hardest challenges for expression-invariant face recognition methods.

In this paper, we employ local features to recognize frontal faces across expressions by applying the recently proposed probabilistic linear discriminant analysis (PLDA) [7]. Local features are more stable against expression

changes than the holistic ones. For example, on smiling faces, only the mouth and the jaw deform significantly from the neutral shape while areas such as eyes, nose, and eyebrows only change a little. For local features, the PLDA has shown very good discrimination even when the features or the fusion rules are simple. This method has also achieved superb performances in neutral face recognition [7] and face recognition across pose [8]. However, to the best of our knowledge, it has never been specifically applied to face identification across expressions. We aim to recognize the faces using one sample image (at least) per subject of the gallery. Therefore, we make a clear distinction between training sets and the gallery. We use the training sets to build a model during the offline phase and these sets contain images of various expressions as well as various subjects. Once this model has been built, it can be used in the online phase to recognize other (test) subjects whose samples are collected in the gallery.

The rest of the paper is organized as follows. Section 2 briefly summarizes the related works while Section 3 describes the PLDA method. Section 4 presents several PLDA approaches to recognize faces across expressions. Cross-database evaluation and comparison to other similar works are discussed in Section 5. Finally, conclusions are drawn in Section 6. More elaborate results regarding recognition rates over various local areas will also be discussed throughout the paper.

II. RELATED WORKS

To cope with facial expressions, several face recognition approaches have been proposed. The first one is morphing probe images to obtain the same expression as of the gallery images [9–11]. With this approach, the lack of texture information (e.g. from closed eyes) can be an issue especially when the features are holistic. The second approach employs dimension reduction methods to suppress the effects of expression differences. The rotated adaptive PCA [12] has been proposed for this purpose. Nevertheless, the method is likely to be suboptimal since the rotation only concentrates on the within-class covariance. The third approach proposes similarity measures which are robust to expression changes. Yu et al. [6] compute motion vectors between faces and learn the within and between-class

variations to derive similarity scores. Those motion vectors though may not be reliable since different persons have their own ways to express a certain expression. Khwaja et al. [13] propose the use of structural similarity index instead of MSE to improve the sparse representation method [14]. Inspired by [14], Nagesh and Baoxin [15] develop a joint sparsity model which sparsely represents the expression-generated variations of an individual and produces robust distance measures. These later methods [13–15] however need samples of various expressions for each subject to be recognized. The other approach assigns weights to various parts of the face based on their stabilities over specific or non-specific expression changes. Martinez [16] uses these weights to compute weighted Euclidean distances between faces while Martinez [5] employs the weights in the fusion of local areas' matching scores.

III. THE PROBABILISTIC LINEAR DISCRIMINANT ANALYSIS

If we denote the j -th observation of the i -th individual as \mathbf{x}_{ij} , then the data generation process of the probabilistic linear discriminant analysis [7] can be expressed as

$$\mathbf{x}_{ij} = \mu + \mathbf{F}\mathbf{h}_i + \mathbf{G}\mathbf{w}_{ij} + \varepsilon_{ij}. \quad (1)$$

Each observed data point \mathbf{x}_{ij} is assumed to be generated from \mathbf{h}_i and \mathbf{w}_{ij} which are points in latent spaces. We call the space of \mathbf{h}_i as the between-individual space and the space of \mathbf{w}_{ij} as the within-individual space. As indicated by the subscripts, observations from the same individual share the same value of \mathbf{h} but have their own values of \mathbf{w} . Hence, the term \mathbf{h} is also called the latent identity variable since it is unique for each individual. The vectors \mathbf{h}_i and \mathbf{w}_{ij} should have smaller lengths than the vector \mathbf{x}_{ij} . They are mapped to the observation space via linear transformations \mathbf{F} and \mathbf{G} respectively and the addition of the observation mean μ and the residual noise ε_{ij} . Note that \mathbf{h} , \mathbf{w} , and ε have multivariate Gaussian distributions and (1) can be described in terms of conditional probabilities:

$$P(\mathbf{x}_{ij}|\mathbf{h}_i, \mathbf{w}_{ij}, \theta) = g_{\mathbf{x}}[\mu + \mathbf{F}\mathbf{h}_i + \mathbf{G}\mathbf{w}_{ij}, \Sigma], \quad (2)$$

$$P(\mathbf{h}_i) = g_{\mathbf{h}}[0, \mathbf{I}], \quad (3)$$

$$P(\mathbf{w}_{ij}) = g_{\mathbf{w}}[0, \mathbf{I}]. \quad (4)$$

Here, $\theta = (\mu, \mathbf{F}, \mathbf{G}, \Sigma)$ is the model parameters and Σ is the diagonal covariance matrix of the residual noise ε . There are two phases in using this model. In the *training (offline) phase*, we learn the parameters $\theta = (\mu, \mathbf{F}, \mathbf{G}, \Sigma)$ using the training data \mathbf{x}_{ij} . In the *recognition (online) phase*, we use the trained model to infer the identities of the probe data.

To learn the model parameters in the training phase, the EM algorithm for factor analysis [17] is employed as follows

(i) **Initialize** $\mathbf{F}, \mathbf{G}, \Sigma$

(ii) **Repeat until** converged

Compute $\mu = (1/N)\sum_{ij}\mathbf{x}_{ij}$ if it is the first iteration

E-step: Compute $E(\mathbf{z}_{ij}|\mathbf{x}_{i\bullet})$ and $E(\mathbf{z}_{ij}\mathbf{z}_{ij}^T|\mathbf{x}_{i\bullet})$ for each data point \mathbf{x}_{ij} , given the current θ

M-step: Compute the new \mathbf{F}, \mathbf{G} , and Σ using the values obtained from the E-step

At the above steps, N is the number of data points and $\mathbf{z}_{ij} = [\mathbf{h}_i^T \mathbf{w}_{ij}^T]^T$. In the PLDA, data points with the same identity are generated from the same \mathbf{h} . Hence, the expectation of \mathbf{z} and $\mathbf{z}\mathbf{z}^T$ are computed simultaneously for all data points of a particular individual. We use the term $\mathbf{x}_{i\bullet}$ to represent all data points of the i -th individual and $E(\mathbf{z}_{ij}|\mathbf{x}_{i\bullet})$ and $E(\mathbf{z}_{ij}\mathbf{z}_{ij}^T|\mathbf{x}_{i\bullet})$ are obtained by computing

$$E(\mathbf{y}|\mathbf{x}_{i\bullet}) = (\mathbf{I} + \mathbf{A}^T \Sigma'^{-1} \mathbf{A})^{-1} \mathbf{A}^T \Sigma'^{-1} (\mathbf{x}' - \mu'), \quad (5)$$

$$E(\mathbf{y}\mathbf{y}^T|\mathbf{x}_{i\bullet}) = (\mathbf{I} + \mathbf{A}^T \Sigma'^{-1} \mathbf{A})^{-1} + E(\mathbf{y}|\mathbf{x}_{i\bullet})E(\mathbf{y}|\mathbf{x}_{i\bullet})^T, \quad (6)$$

where $\mathbf{y} = [\mathbf{h}_i^T \mathbf{w}_{i1}^T \mathbf{w}_{i2}^T \dots]^T$ and extracting $E(\mathbf{z}_{ij}|\mathbf{x}_{i\bullet})$ and $E(\mathbf{z}_{ij}\mathbf{z}_{ij}^T|\mathbf{x}_{i\bullet})$ from $E(\mathbf{y}|\mathbf{x}_{i\bullet})$ and $E(\mathbf{y}\mathbf{y}^T|\mathbf{x}_{i\bullet})$ accordingly. The matrix \mathbf{A} , \mathbf{x}' , and μ' are formed by arranging the generative equations for all $\mathbf{x}_{i\bullet}$

$$\begin{bmatrix} \mathbf{x}_{i1} \\ \mathbf{x}_{i2} \\ \vdots \end{bmatrix} = \begin{bmatrix} \mu \\ \mu \\ \vdots \end{bmatrix} + \begin{bmatrix} \mathbf{F} & \mathbf{G} & \mathbf{0} & \dots & \mathbf{0} \\ \mathbf{F} & \mathbf{0} & \mathbf{G} & \dots & \mathbf{0} \\ \vdots & \vdots & \vdots & \dots & \vdots \\ \mathbf{F} & \mathbf{0} & \mathbf{0} & \dots & \mathbf{G} \end{bmatrix} \begin{bmatrix} \mathbf{h}_i \\ \mathbf{w}_{i1} \\ \mathbf{w}_{i2} \\ \vdots \end{bmatrix} + \begin{bmatrix} \varepsilon_{i1} \\ \varepsilon_{i2} \\ \vdots \end{bmatrix}, \quad (7)$$

$$\mathbf{x}' = \mu' + \mathbf{A}\mathbf{y} + \varepsilon', \quad (8)$$

and Σ' is the diagonal covariance matrix of ε' . The new \mathbf{F}, \mathbf{G} , and Σ are then computed as follows

$$[\mathbf{F}_{new} \mathbf{G}_{new}] = (\sum_{ij}\mathbf{x}_{ij}E(\mathbf{z}_{ij}|\mathbf{x}_{i\bullet}))(\sum_{ij}E(\mathbf{z}_{ij}\mathbf{z}_{ij}^T|\mathbf{x}_{i\bullet}))^{-1}, \quad (9)$$

$$\Sigma_{new} = (1/N)\text{diag}\{\sum_{ij}(\mathbf{x}_{ij}\mathbf{x}_{ij}^T - [\mathbf{F}_{new} \mathbf{G}_{new}]E(\mathbf{z}_{ij}|\mathbf{x}_{i\bullet})\mathbf{x}_{ij}^T)\}. \quad (10)$$

To perform closed-set identification in the recognition phase, a model comparison approach has been developed in [7] to compute the likelihood that particular probe and gallery data are matched i.e. generated from the same \mathbf{h} . If we denote the probe image as \mathbf{x}^p and the gallery images as $\mathbf{x}_1 \dots \mathbf{x}_M$, then the likelihood that \mathbf{x}^p matches one of the gallery images \mathbf{x}_m is defined as

$$P(\mathbf{x}_1 \dots \mathbf{x}_M, \mathbf{x}^p | \mathbf{M}_m) = P(\mathbf{x}^p, \mathbf{x}_m) \prod_{n=1 \dots M, n \neq m} P(\mathbf{x}_n). \quad (11)$$

\mathbf{M}_m represents the situation that \mathbf{x}^p matches \mathbf{x}_m but not the other gallery images. $P(\mathbf{x}^p, \mathbf{x}_m)$ and $P(\mathbf{x}_n)$ can be obtained by arranging the generative equations for $\mathbf{x}^p, \mathbf{x}_m$ and \mathbf{x}_n as (7) respectively and compute $g_{\mathbf{x}}[\mu', \mathbf{A}\mathbf{A}^T + \Sigma']$. The matched identity can be chosen as the m which the $P(\mathbf{x}_1 \dots \mathbf{x}_M, \mathbf{x}^p | \mathbf{M}_m)$ is the maximum.

IV. THE PLDA APPROACHES FOR FACE RECOGNITION ACROSS EXPRESSIONS

In this section, we consider several approaches in using the probabilistic linear discriminant analysis for recognition across expressions. These approaches are evaluated on the same data so that the results can be compared.

A. The basic PLDA approach

With the basic PLDA approach, we actually apply the method explained in Section 3. This approach is somewhat similar to the ones proposed in [5, 16] in the sense that pixels and local areas are assigned different weights and the recognition is performed through local matching.

We use the Cohn-Kanade (CK) database [18] to construct the training and the test data. This database contains 97 subjects showing different expressions. There are seven expression labels: neutral, angry, disgust, feared, happy, sad, and surprised. From the database, we choose 49 subjects from each of which we collect 4 images of each expression label and include them in the training data. From each of the remaining 48 subjects, we choose 1 neutral image to be used as the gallery image and 4 images per non-neutral expression label to be included in the probe data. Note that all non-neutral images are of the peak expressions.

From the training and the test images, we extract local features from 29 key points on the face (Fig. 1). Locations of these key points are obtained from the manual annotations provided in [19]. After an image is rescaled to have 128-pixels eye-to-eye distance, an 81×81 patch is extracted from every key point and converted to a feature vector. All feature vectors consist of grayscale values which are normalized to have zero mean and variance 1. From this procedure we get 29 training sets, 29 probe sets, and 29 galleries each of which corresponds to a particular key point.

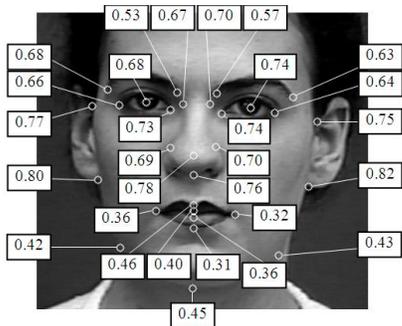


Figure 1. The first-match correct rates of the individual PLDA models which correspond to 29 predefined key points.

In the training phase, we construct a PLDA model from each training set. We include 16 basis vectors in \mathbf{F} and 16 basis vectors in \mathbf{G} . We then test the recognition performance of each PLDA model individually i.e. inferring the identity based on single local area only. Fig. 1 shows the first-match correct rates of the individual models connected to the corresponding key points. It can be seen that areas around the mouth, jaw, and chin are the least discriminative (0.31 –

0.46 correct rates) while areas around the cheeks are the most discriminative (0.75 – 0.82 correct rates).

We also fuse the local matching scores (likelihoods) by multiplying them throughout the 29 local areas as well as a number of their subsets. This way we treat the likelihoods to be independent each other. Table 1 presents the evaluation results after the fusion. It clearly shows that the fusion significantly improves correct rates indicating that information from different local areas complements each other. We find again here that upper regions such as eyebrows, eyes, nose, and cheeks are more discriminative than lower regions such as mouth, chin, and jaw. The high correct rates also suggested that the basic PLDA approach is robust to expressions at least for frontal face images.

TABLE I. FIRST-MATCH CORRECT RATES OF THE BASIC PLDA APPROACH WITH FUSION ON DIFFERENT SUBSETS OF KEY POINTS

Subset (Points on the)	First-Match Correct Rate	Subset (Points on the)	First-Match Correct Rate
all points	0.999	nose	0.958
cheeks	0.973	mouth	0.730
eye+eyebrows	0.987	jaw+chin	0.626

In the previous experiment, we register face images based on two points on the eyes i.e. rescaling the images to have 128-pixels eye-to-eye distance. This kind of registration method is quite simple and retains the expressions of the faces. The other way of doing the registration is by applying piece-wise triangular warp as has been done in [9, 10]. This registration method convert face shapes into a standard shape thereby removing the expressions posed by the faces. In the following experiment, we repeat the whole process of the previous experiment except we apply the piece-wise triangular warp before extracting local features. We use a 228×208 standard face shape and 59 points provided in the manual annotations of [19] to warp the images (Fig. 2). The first-match correct rates of the individual PLDA models as well as of the fused models are presented in Fig. 3 and Table 2. Compared to the previous results, significant improvements are obtained in all individual models and the fusion over the mouth, jaw, and chin regions. From this point onward, all experiments will employ piece-wise triangular warp to perform the registration.

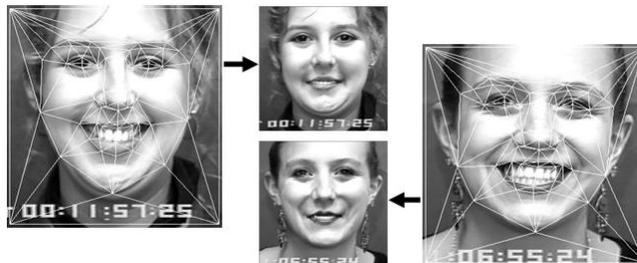


Figure 2. Face images are warped to a standard shape using piece-wise triangular warp based on 59 points on the faces.

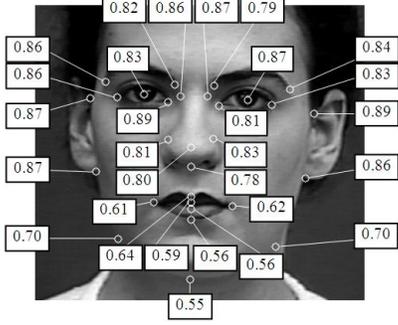


Figure 3. The first-match correct rates of the individual PLDA models when the piece-wise triangular warp is applied.

TABLE II. FIRST-MATCH CORRECT RATES OF THE BASIC PLDA APPROACH WITH SCORE FUSION AND PIECE-WISE TRIANGULAR WARP.

Subset (Points on the)	First-Match Correct Rate	Subset (Points on the)	First-Match Correct Rate
all points	0.995	nose	0.957
cheeks	0.977	mouth	0.862
eye+eyebrows	0.977	jaw+chin	0.872

B. PLDA approaches which utilize expression labels

Face images are sometimes enclosed by additional information such as pose or expression labels which are acquired manually or automatically in advance. Martinez [5] has suggested that expression labels can be used to improve recognition performances. In his work, he set the weight of a particular local area to vary from expression to expression. In this subsection, we try to employ expression labels using two PLDA approaches: expression-specific PLDA and tied PLDA. The later approach has been successfully applied to face recognition across pose [8].

With the expression-specific PLDA, for each key point, we construct six PLDA models each of which is trained using data with neutral and one particular non-neutral label. Hence, from each training set of the previous experiment, we construct six training subsets: “neutral-angry” subset, “neutral-disgust” subset, and so on. The constructed models are then called “neutral-angry” model, “neutral-disgust” model, and so on. In the recognition phase, we compute local matching likelihoods using the appropriate models. For example, if the probe image has an “angry” label, then we use the “neutral-angry” model to obtain the likelihoods (note that all gallery images have “neutral” labels).

The second approach for making use of the expression labels is the tied PLDA [8]. Again for each key point we construct six models: “neutral-angry” model, “neutral-disgust” model, and so on using the same training subsets as before. Each model however has two parameter sets $\theta_1 = (\mu_1, \mathbf{F}_1, \mathbf{G}_1, \Sigma_1)$ and $\theta_2 = (\mu_2, \mathbf{F}_2, \mathbf{G}_2, \Sigma_2)$ which correspond to the neutral and non-neutral labels respectively. If we denote the training data as \mathbf{x}_{ijk} with $k = 1, 2$ represents the expression labels, then the model can be trained using the EM algorithm described in Section 3 but two parameter sets are initialized

and iteratively updated. In each iteration, $\theta_k = (\mu_k, \mathbf{F}_k, \mathbf{G}_k, \Sigma_k)$ are updated using $\mathbf{x}_{\bullet\bullet k}$ and the corresponding $E(\mathbf{z}_{ijk}|\mathbf{x}_{i\bullet\bullet})$ and $E(\mathbf{z}_{ijk}\mathbf{z}_{ijk}^T|\mathbf{x}_{i\bullet\bullet})$. Here, $\mathbf{z}_{ijk} = [\mathbf{h}_i^T \mathbf{w}_{ijk}^T]^T$ and again $E(\mathbf{z}_{ijk}|\mathbf{x}_{i\bullet\bullet})$ and $E(\mathbf{z}_{ijk}\mathbf{z}_{ijk}^T|\mathbf{x}_{i\bullet\bullet})$ are derived simultaneously for all $\mathbf{x}_{i\bullet\bullet}$ by using generative equations as follows

$$\begin{bmatrix} \mathbf{x}_{i11} \\ \mathbf{x}_{i21} \\ \vdots \\ \mathbf{x}_{i12} \\ \mathbf{x}_{i22} \\ \vdots \end{bmatrix} = \begin{bmatrix} \mu_1 \\ \mu_1 \\ \vdots \\ \mu_2 \\ \mu_2 \\ \vdots \end{bmatrix} + \begin{bmatrix} \mathbf{F}_1 & \mathbf{G}_1 & \mathbf{0} & \dots & \mathbf{0} & \mathbf{0} & \dots \\ \mathbf{F}_1 & \mathbf{0} & \mathbf{G}_1 & \dots & \mathbf{0} & \mathbf{0} & \dots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \mathbf{F}_2 & \mathbf{0} & \mathbf{0} & \dots & \mathbf{G}_2 & \mathbf{0} & \dots \\ \mathbf{F}_2 & \mathbf{0} & \mathbf{0} & \dots & \mathbf{0} & \mathbf{G}_2 & \dots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \end{bmatrix} \begin{bmatrix} \mathbf{h}_i \\ \mathbf{w}_{i11} \\ \mathbf{w}_{i21} \\ \vdots \\ \mathbf{w}_{i12} \\ \mathbf{w}_{i22} \\ \vdots \end{bmatrix} + \begin{bmatrix} \varepsilon_{i11} \\ \varepsilon_{i21} \\ \vdots \\ \varepsilon_{i12} \\ \varepsilon_{i22} \\ \vdots \end{bmatrix}. \quad (12)$$

In the recognition phase, the appropriate tied models are picked and the likelihoods are computed based on (12).

The evaluation results of the two approaches over the CK database are shown in Table 3 and Table 4. Compared to the previous approach, the expression-specific PLDA clearly has better performances in all fusion cases suggesting that multi-modal distributions may describe the data better. The tied PLDA on the other hand has worse performances on the fusion over the nose, mouth, jaw, and chin regions. This indicates that unlike data with varying pose, data with varying expressions are not well separated when they are “semantically” grouped based on the expression labels.

TABLE III. FIRST-MATCH CORRECT RATES OF THE EXPRESSION-SPECIFIC PLDA APPROACH WITH FUSION ON DIFFERENT SUBSETS OF KEY POINTS.

Subset (Points on the)	First-Match Correct Rate	Subset (Points on the)	First-Match Correct Rate
all points	0.996	nose	0.966
cheeks	0.995	mouth	0.881
eye+eyebrows	0.991	jaw+chin	0.894

TABLE IV. FIRST-MATCH CORRECT RATES OF THE TIED PLDA APPROACH WITH FUSION ON DIFFERENT SUBSETS OF KEY POINTS

Subset (Points on the)	First-Match Correct Rate	Subset (Points on the)	First-Match Correct Rate
all points	0.995	nose	0.929
cheeks	0.994	mouth	0.680
eye+eyebrows	0.980	jaw+chin	0.857

C. The mixture of PLDAs approach

From the previous experiments, we find that multi-modal distributions might improve recognition performances. In this subsection, we employ a mixture of PLDA models to multi-modally describe the data. This somehow emulates the

expression-specific PLDA without requiring the data to be labeled.

The mixture of PLDAs approach can be described using (2) – (4) as in Section 3 but (2) is reformulated as

$$P(\mathbf{x}_{ij}|\mathbf{h}_{i1 \dots iK}, \mathbf{w}_{ij1 \dots ijK}, \theta_1 \dots \theta_K, \pi_1 \dots \pi_K) = \sum_{k=1 \dots K} \pi_k g_{\mathbf{x}}[\mu_k + \mathbf{F}_k \mathbf{h}_{ik} + \mathbf{G}_k \mathbf{w}_{ijk}, \Sigma_k]. \quad (13)$$

We now have K parameter sets (θ_k, π_k) , $k = 1 \dots K$ each of which corresponds to a mixture's component. The term $\theta_k = (\mu_k, \mathbf{F}_k, \mathbf{G}_k, \Sigma_k)$ has the same interpretation as before and π_k is the prior probability of a measurement belonging to the k -th component.

To train the mixture of PLDA models, we follow the receipt described in [20] and come up with the EM algorithm presented below

- (i) **Initialize** $\pi_k, \mu_k, \mathbf{F}_k, \mathbf{G}_k, \Sigma_k$, $k = 1 \dots K$
- (ii) **Repeat until** converged
 - Compute** $h_{ijk} = E(\omega_k|\mathbf{x}_{ij})$, $k = 1 \dots K$ for all data point \mathbf{x}_{ij}
 - Compute** the new $\mu_k = (1/\sum_{ij} h_{ijk}) \sum_{ij} h_{ijk} \mathbf{x}_{ij}$ and the new $\pi_k = (1/N) \sum_{ij} h_{ijk}$
 - E-step:** Compute $E(\mathbf{z}_{ij}|\mathbf{x}_{i\bullet}, \omega_k)$ and $E(\mathbf{z}_{ij} \mathbf{z}_{ij}^T | \mathbf{x}_{i\bullet}, \omega_k)$, $k = 1 \dots K$ for each data point \mathbf{x}_{ij} , given the current θ_k , $k = 1 \dots K$
 - M-step:** Compute the new $\mathbf{F}_k, \mathbf{G}_k$, and Σ_k , $k = 1 \dots K$ using the values obtained from the E-step

The term ω_k represents the k -th mixture component and

$$h_{ijk} = E(\omega_k|\mathbf{x}_{ij}) \propto P(\mathbf{x}_{ij}, \omega_k) = \pi_k g_{\mathbf{x}}[\mu_k, [\mathbf{F}_k \mathbf{G}_k][\mathbf{F}_k \mathbf{G}_k]^T + \Sigma_k]. \quad (14)$$

To update $\mathbf{F}_k, \mathbf{G}_k$, and Σ_k we use the following formulas

$$[\mathbf{F}_{k-new} \mathbf{G}_{k-new}] = (\sum_{ij} h_{ijk} \mathbf{x}_{ij} E(\mathbf{z}_{ij}|\mathbf{x}_{i\bullet}, \omega_k)) (\sum_{ij} h_{ijk} E(\mathbf{z}_{ij} \mathbf{z}_{ij}^T | \mathbf{x}_{i\bullet}, \omega_k))^{-1}, \quad (15)$$

$$\Sigma_{k-new} = (1/N) \text{diag} \{ \sum_{ij} h_{ijk} (\mathbf{x}_{ij} \mathbf{x}_{ij}^T - [\mathbf{F}_{k-new} \mathbf{G}_{k-new}] E(\mathbf{z}_{ij}|\mathbf{x}_{i\bullet}, \omega_k) \mathbf{x}_{ij}^T) \}. \quad (16)$$

In the recognition phase, we compute $\sum_{k=1 \dots K} \pi_k g_{\mathbf{x}}[\mu_k', \mathbf{A}_k \mathbf{A}_k^T + \Sigma_k]$ analogously to the basic PLDA approach to obtain $P(\mathbf{x}_n)$, $P(\mathbf{x}^p, \mathbf{x}_m)$, and the matching likelihoods.

We apply the mixture of PLDAs approach with 6 components to the same training and test data as the first experiment. The evaluation results are presented in Table 5. It can be seen that the mixture of PLDAs has a bit lower performances than the expression-specific PLDA when the fusion is performed over the cheeks, eyes, and eyebrows. Further performance inferiorities appear on the fusion over the mouth. Nevertheless, when the fusion is performed over the jaw, chin, and the whole facial regions, the mixture of PLDAs outperforms the expression-specific PLDA.

TABLE V. FIRST-MATCH CORRECT RATES OF THE MIXTURE OF PLDAS APPROACH WITH FUSION ON DIFFERENT SUBSETS OF KEY POINTS.

Subset (Points on the)	First-Match Correct Rate	Subset (Points on the)	First-Match Correct Rate
all points	1.000	nose	0.954
cheeks	0.992	mouth	0.828
eye+eyebrows	0.990	jaw+chin	0.913

V. CROSS-DATABASE EVALUATION AND COMPARISON TO OTHER SIMILAR WORKS

To further confirm the robustness of the PLDA approaches against expression variations, we conduct another evaluation on the JAFFE database [21]. This database contains 10 female subjects showing expressions with seven labels as in the CK database. We use all the 10 subjects to create the gallery and the probe data. From each subject, we choose 1 neutral image to be used as the gallery image and 3 images per non-neutral expression label to be included in the probe data. This time we do not perform training instead we use the models previously trained on the CK database. To locate the 29 key points on the faces, we use an active appearance model [22]. Table 6 presents the evaluation performances of the four PLDA approaches and the highest score is achieved by the mixture of PLDAs. The correct rates are quite high which means that in addition to expression-robust the PLDA models also have good generalizations.

TABLE VI. FIRST-MATCH CORRECT RATES OF SEVERAL PLDA APPROACHES ON THE JAFFE DATABASE.

Approach	First-Match Correct Rate	Approach	First-Match Correct Rate
Basic PLDA	0.960	Tied PLDA	0.943
Expression-specific PLDA	0.949	Mixture of PLDAs	0.966

We also compare the performances of the PLDA approaches to other similar works which use the same datasets as this paper. As shown in Table 7, the mixture of PLDAs clearly outperforms the weighted Euclidean distance method proposed in [16]. It also demonstrates similar performances to the B-JSM method proposed in [15]. However, always keep in mind that the mixture of PLDAs only uses one gallery image per subject while the B-JSM method requires multiple images of various expressions.

VI. CONCLUSION

This paper applies the probabilistic linear discriminant analysis (PLDA) to face recognition across expressions. Probe images with non-neutral expressions are matched against neutral images of the gallery. We test four PLDA approaches on the Cohn-Kanade (CK) and JAFFE databases and obtain high correct rates (0.943 – 1.000 for the fusion over all facial regions). This indicates that the PLDA

approaches are robust to expression variations. In addition to that, good results of the cross-database (CK and JAFFE) evaluation suggested that the approaches have good generalizations. Among the approaches, the mixture of PLDAs achieves the highest correct rates. The experiments also indicate that areas around the cheeks, eyes, and eyebrows are more discriminative than areas around the mouth, jaw, chin, and nose.

TABLE VII. PERFORMACE COMPARISON OF THE PLDA APPROACHES TO OTHER SIMILAR WORKS.

Reference	Data-base	Num. of Test Subjects	Num. of Gallery Images per Subject	Correct Rate
Mixture of PLDAs (ours)	JAFFE	10	1	0.966
Weighted Euclidean [16]	JAFFE	10	1	± 0.75
B-JSM [15]	JAFFE	10	4	0.961
Mixture of PLDAs (ours)	CK	48	1	1.000
B-JSM [15]	CK	97	10	0.971

REFERENCES

- [1] M. Turk and A. Pentland, "Eigenfaces for recognition," *J. Cognitive Neuroscience*, vol. 3, pp. 71-86, 1991.
- [2] P. N. Belhumeur, J. P. Hespanha, and D. J. Kriegman, "Eigenfaces vs. fisherfaces: Recognition using class specific linear projection," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 19, pp. 711-720, 1997.
- [3] W. Xiaogang and T. Xiaoou, "Dual-space linear discriminant analysis for face recognition," *Proc. IEEE Comput. Soc. Conf. Computer Vision and Pattern Recognition*, vol. 2, 2004, pp. II-564-II-569.
- [4] L. Wiskott, J. M. Fellous, N. Kuiger, and C. von der Malsburg, "Face recognition by elastic bunch graph matching," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 19, pp. 775-779, 1997.
- [5] A. M. Martinez, "Recognizing imprecisely localized, partially occluded, and expression variant faces from a single sample per class," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 24, pp. 748-763, 2002.
- [6] B. Yu, P. Chen, and L. Jin, "Recognizing faces with expressions: Within-class space and between-class space," *Proc. 16th Int. Conf. Pattern Recognition*, vol. 1, 2002, pp. 139-142.
- [7] S. J. D. Prince and J. H. Elder, "Probabilistic linear discriminant analysis for inferences about identity," *Proc. 11th IEEE Int. Conf. Computer Vision*, 2007, pp. 1-8.
- [8] S. Prince, P. Li, Y. Fu, U. Mohammed, and J. Elder, "Probabilistic models for inference about identity," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 34, in press.
- [9] M. Ramachandran, S. K. Zhou, D. Jhalani, and R. Chellappa, "A method for converting a smiling face to a neutral face with applications to face recognition," *Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Processing*, vol. 2, 2005, pp. ii/977-ii/980.
- [10] L. Xiaoxing, G. Mori, and Z. Hao, "Expression-invariant face recognition with expression classification," *Proc. 3rd Canadian Conf. Computer and Robot Vision*, 2006, pp. 77-84.
- [11] H. Chao-Kuei, L. Shang-Hong, and C. Yung-Chang, "Expression-invariant face recognition with constrained optical flow warping," *IEEE Trans. Multimedia*, vol. 11, pp. 600-610, 2009.
- [12] S. Chen and B. C. Lovell, "Illumination and expression invariant face recognition with one sample image," *Proc. 17th Int. Conf. Pattern Recognition*, vol. 1, 2004, pp. 300-303.
- [13] A. Khwaja, A. Asthana, and R. Goecke, "Illumination and expression invariant recognition using SSIM based sparse representation," *Proc. 20th Int. Conf. Pattern Recognition*, 2010, pp. 4028-4031.
- [14] J. Wright, A. Y. Yang, A. Ganesh, S. S. Sastry, and M. Yi, "Robust face recognition via sparse representation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 31, pp. 210-227, 2009.
- [15] P. Nagesh and L. Baoxin, "A compressive sensing approach for expression-invariant face recognition," *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, 2009, pp. 1518-1525.
- [16] A. M. Martinez, "Recognizing expression variant faces from a single sample image per class," *Proc. 2003 IEEE Comput. Soc. Conf. Computer Vision and Pattern Recognition*, vol. 1, 2003, pp. I-353-I-358.
- [17] D. B. Rubin and D. T. Thayer, "EM algorithms for ML factor analysis," *Psychometrika*, vol. 47, pp. 69-76, 1982.
- [18] T. Kanade, J. F. Cohn, and T. Yingli, "Comprehensive database for facial expression analysis," *Proc. Fourth IEEE Int. Conf. Automatic Face and Gesture Recognition*, 2000, pp. 46-53.
- [19] G. Lipori, "Manual annotations of facial fiducial points on the Cohn Kanade database," LAIV Laboratory, University of Milan, web url: <http://lipori.dsi.unimi.it/download.html>.
- [20] Z. Ghahramani and G. E. Hinton, "The EM algorithm for mixtures of factor analyzers," University of Toronto, Tech. Rep., 1996.
- [21] M. Lyons, S. Akamatsu, M. Kamachi, and J. Gyoba, "Coding facial expressions with Gabor wavelets," *Proc. Third IEEE Int. Conf. Automatic Face and Gesture Recognition*, 1998, pp. 200-205.
- [22] I. Matthews and S. Baker, "Active appearance models revisited," *Int. J. Computer Vision*, vol. 60, pp. 135-164, 2004.