

Emotion recognition from facial images with arbitrary views

Xiaohua Huang
huang.xiaohua@ee.oulu.fi

Guoying Zhao
gyzhao@ee.oulu.fi

Matti Pietikäinen
mkp@ee.oulu.fi

Center for Machine Vision Research
Department of Computer Science and
Engineering
University of Oulu
Oulu, Finland

Abstract

Facial expression recognition has been predominantly utilized to analyze the emotional status of human beings. In practice nearly frontal-view facial images may not be available. Therefore, a desirable property of facial expression recognition would allow the user to have any head pose. Some methods on non-frontal-view facial images were recently proposed to recognize the facial expressions by building discriminative subspace in specific views. We argue that this kind of approach ignores (1) the discrimination of inter-class samples with the same view label and (2) the closeness of intra-class samples with all view labels. This paper proposes a new method to recognize arbitrary-view facial expressions by using discriminative neighborhood preserving embedding and multi-view concepts. It first captures the discriminative property of inter-class samples. In addition, it explores the closeness of intra-class samples with arbitrary view in a low-dimensional subspace. Experimental results on BU-3DFE and Multi-PIE databases show that our approach achieves promising results for recognizing facial expressions with arbitrary views.

1 Introduction

A goal of automatic facial expression analysis is to determine the emotional state, e.g. happiness, sadness and so on, of human beings based on facial images [20]. Despite the highly developed ability of humans to obtain information from visual observation of faces, facial expression recognition remains a very challenging task for computer vision. This challenge is specially noticeable when frontal-view facial images are not available. With some multi-view databases available [4, 19], the view-invariant approaches have been under investigation by the research community in facial expression recognition. These approaches can be categorized into two groups: (1) one considers facial expressions separated from view labels, while (2) the other couples facial expression labels with view labels in order to reduce the influence of pose variance.

The view-dependent approach is a general framework to recognize facial expressions on arbitrary views [7, 8, 13, 15, 16]. It consists of 2D pose estimation and emotion classification. In [7, 8] Hu et al. proposed view-dependent emotion classifiers to investigate facial

expression recognition problem with five yaw views. In addition, they used this system to evaluate various feature descriptors and dimensionality reduction methods. In [13], Moore and Bowden further used this framework to evaluate local binary patterns and its variations for recognizing facial expressions. Recently, in [15] Rudovic et al. presented to recognize facial expressions using 2-D geometric features and a regression-based scheme, while it requires that the pose and location of facial landmarks for each image is prior known. In [16], they furthermore proposed to use a Gaussian process regression model for mapping the 2D locations of landmark points of facial expressions in non-frontal views to the corresponding locations in frontal view.

From these works [7, 8, 13], pose estimation may be viewed as a critical element, although multi-view 2D tracking methods [2, 24] could be used to register facial images, yet still few studies use 2D tracking techniques for multi-view facial expression recognition. Whatever these works [7, 8, 13] based on pose estimation and emotion recognition achieved good performance, they still had to build separate subspaces or regression models from facial images with specific view labels. This is inevitable to in advance estimate the head pose of any probe, while pose estimation is still a challenging problem in computer vision.

For resisting limitations of the view-dependent method, some researchers take facial expression labels and view labels together into consideration. In [22], Zheng et al. proposed the unified Bayes theoretical framework to classify facial expressions, where the recognition problem can be formulated as an optimization problem of minimizing an upper bound of Bayes error. They consider the facial images from the same view and facial expression label as an independent subclass. But among their method, they used SIFT features extracted at 83 landmark points from the original 3D face model to represent facial images. This may severely limit their practical application while no 3D face model is available. Instead of 2-D geometric landmarks, in [23] Zheng et al. presented to use the regional covariance matrix representation for appearance features. Additionally, they proposed the Bayes discriminant analysis via GMM (BDA/GMM) to reduce the dimensionality of feature vectors while preserving the most discriminative information.

Comparing with the view-dependent facial expression recognition methods that require to estimate the view angle accurately, the approaches [22, 23] explored the correlation of different samples by coupling facial expressions with views. Their works showed that the samples with the same expression label may be similar even in different view angles. Motivated partly by their work, our paper will address the following questions:

(1) *Whether there exists discriminative features for facial expressions or not?* Recent research [7, 13, 23] shows appearance-based features achieve a good performance, while they may carry irrelevant information of face identity. This irrelevant information can confuse the classifier for recognizing facial expressions. Therefore, it is necessary to enhance the discriminative power of the features.

(2) *How can we utilize the correlation between facial expressions and views?* The traditional framework [7, 8] regards the view and facial expression as two separate information. It usually requires the great achievement of pose estimation. It therefore leads a challenge to this framework. To address this, the exploration of their correlation may be a good way to make this system simple and avoid from the cumulative error caused by pose estimation.

Contribution. This paper introduces a novel view-independent framework to recognize facial expression on arbitrary views. In this framework, we propose discriminative neighbor preserving embedding using neighbor graph and maximizing margin criterion. Additionally, we propose to use multi-view framework [1] and multiset canonical correlation analysis (MCCA) [14] for coupling facial expressions with views.

To explain the concepts in our approach, the paper has been organized as follows. In Section 2, we explain our proposed approach for exploring the discriminative information of facial expressions independent to the views. The results of applying the proposed method for recognizing facial expressions are provided in Section 3. Finally we summarize our findings in Section 4.

2 Proposed methodology

In this section, we present a new multi-view discriminative framework using MCCA [14] and multi-view model theorem [1] for recognizing facial expressions with arbitrary views. The algorithm has the following characteristics: (1) it respects the intrinsic and discriminant structure of samples, (2) it effectively preserves the local manifold structure within each class and meanwhile enlarges the margins between different classes, and (3) it regards emotion recognition from facial images with arbitrary views as a multi-view optimization problem.

2.1 Discriminative Neighbor Preserving Embedding (DNPE)

Dimensionality reduction methods, especially manifold learning methods, have been widely used to obtain discriminative information for facial expression recognition, e.g. locality preserving projection (LPP) [5, 17] and neighbor preserving embedding (NPE) [6]. Linearized manifold learning methods aim to preserve the local manifold structure of samples, in which the k-nearest neighbor searching method is often used to realize the local manifold structure. However, the unsupervised nature leads to the loss of the discriminative capability. Instead, the supervised intra-class structure explores their discriminative capability using class information [11], but it is not sufficient to represent the boundary of inter-class samples. In [12], the marginal neighboring graph was recently proposed to describe the pairwise inter-class boundaries.

Given n training images with C classes, they are denoted as $\mathbf{X} = [\vec{x}_1, \dots, \vec{x}_n] \in R^{D \times n}$, where D is the dimensionality of the feature vector. With the class label and Euclidean distance, we construct the within-class set Ω_p^{wi} ('wi' means within-class) of the sample \vec{x}_p , where $p = 1, \dots, n$. This set contains k^{wi} nearest neighbors \vec{x}_q ($q = 1, \dots, k^{\text{wi}}$) of \vec{x}_p with the same label of \vec{x}_p . Motivated by within-class concept in linear discriminant analysis (LDA) [3], we first design an intrinsic graph $\mathbf{G}^{\text{wi}} = \{\mathbf{X}, \mathbf{V}^{\text{wi}}\}$ that preserves the intrinsic structure of intra-class samples, where \mathbf{V}^{wi} is the similarity matrix of \mathbf{G}^{wi} .

In the intrinsic graph \mathbf{G}^{wi} , we aim to preserve the similarity between intra-class vertex pairs. It is noted that any sample \vec{x}_p can be linearly reconstructed by its intra-class samples \vec{x}_q . Therefore there exists the reconstruction weight matrix \mathbf{V}^{wi} for \mathbf{X} that can be obtained by minimizing the following formulation,

$$\mathcal{E}^{\text{wi}}(\mathbf{X}) = \sum_p \left\| \vec{x}_p - \sum_{\vec{x}_q \in \Omega_p^{\text{wi}}} v_{p,q}^{\text{wi}} \vec{x}_q \right\|^2, \quad (1)$$

with the constraint $\sum_q v_{p,q}^{\text{wi}} = 1$, where $v_{p,q}^{\text{wi}}$ in \mathbf{V}^{wi} is the weight of the edge from \vec{x}_p to \vec{x}_q .

Given the lower dimensional feature space $\mathbf{U} \in R^{d \times d}$, where d ($d \ll D$) is the dimensionality of this space, the sample \vec{x}_p is transformed to this space via $\vec{y}_p = \mathbf{U}^T \vec{x}_p$. Therefore, the sample \vec{y}_p can be represented as a linear combination of its neighbors with the corresponding

coefficients v_{pq}^{wi} . The corresponding cost function is defined as follows,

$$\begin{aligned}\varepsilon^{\text{wi}}(\mathbf{Y}) &= \sum_p \left\| \vec{y}_p - \sum_{\vec{y}_q \in \Omega_p^{\text{wi}}} v_{p,q}^{\text{wi}} \vec{y}_q \right\|^2 \\ &= (\mathbf{y} - \mathbf{V}^{\text{wi}} \mathbf{y})(\mathbf{y} - \mathbf{V}^{\text{wi}} \mathbf{y})^T \\ &= \mathbf{U}^T \mathbf{S}^{\text{wi}} \mathbf{U},\end{aligned}\quad (2)$$

where $\mathbf{S}^{\text{wi}} = \mathbf{X}(\mathbf{I} - \mathbf{V}^{\text{wi}})(\mathbf{I} - \mathbf{V}^{\text{wi}})^T \mathbf{X}^T$ represents the local geometric structure of intra-class samples. Here, we fix the weight matrix while optimizing \mathbf{U} for minimizing Eqn.2.

Following the aforementioned procedure, we also build the between-class set Ω_p^{bw} ('bw' means between-class). It contains k^{bw} samples \vec{x}_r ($r = 1, \dots, k^{\text{bw}}$) nearest neighboring to \vec{x}_p that have different class labels to \vec{x}_p . We further design a penalty graph $\mathbf{G}^{\text{bw}} = \{\mathbf{X}, \mathbf{V}^{\text{bw}}\}$ that describes the margin across inter-class boundaries in the same way to building an intrinsic graph. Among the penalty graph \mathbf{G}^{bw} , the penalty similarity matrix \mathbf{V}^{bw} describes the similarity among \vec{x}_p and inter-class ones \vec{x}_r . It can be obtained by minimizing the following formulation,

$$\varepsilon^{\text{bw}}(\mathbf{X}) = \sum_p \left\| \vec{x}_p - \sum_{\vec{x}_r \in \Omega_p^{\text{bw}}} v_{p,r}^{\text{bw}} \vec{x}_r \right\|^2, \quad (3)$$

with the constraint $\sum_r v_{p,r}^{\text{bw}} = 1$, where $v_{p,r}^{\text{bw}}$ in \mathbf{V}^{bw} represents the weight of the edge from \vec{x}_p to \vec{x}_r with different class labels.

In the lower dimensional feature space, we hopefully maximize the boundary of samples with different class labels. Therefore, \mathbf{U} makes the following objective function maximizing while \mathbf{V}^{bw} is fixed,

$$\begin{aligned}\varepsilon^{\text{bw}}(\mathbf{Y}) &= \sum_p \left\| \vec{y}_p - \sum_{\vec{y}_r \in \Omega_p^{\text{bw}}} v_{p,r}^{\text{bw}} \vec{y}_r \right\|^2 \\ &= (\mathbf{y} - \mathbf{V}^{\text{bw}} \mathbf{y})(\mathbf{y} - \mathbf{V}^{\text{bw}} \mathbf{y})^T \\ &= \mathbf{U}^T \mathbf{S}^{\text{bw}} \mathbf{U},\end{aligned}\quad (4)$$

where $\mathbf{S}^{\text{bw}} = \mathbf{X}(\mathbf{I} - \mathbf{V}^{\text{bw}})(\mathbf{I} - \mathbf{V}^{\text{bw}})^T \mathbf{X}^T$ represents the local geometric structure of inter-class samples.

Maximum margin criterion (MMC) was proposed in [10] to maximize the margin between classes. It was shown that MMC can represent class separability better than principal component analysis (PCA) [18]. Additionally, MMC does not suffer from the small sample size problem. Therefore, our objective function is defined as,

$$\mathbf{U} = \arg \max \{ \mathbf{U}^T \mathbf{S}^{\text{bw}} \mathbf{U} - \eta \mathbf{U}^T \mathbf{S}^{\text{wi}} \mathbf{U} \}, \quad (5)$$

where η is the balancing factor which adjusts the second term to ensure a positive objective function. We empirically observe that η has not much impact on the performance. For avoiding the degeneration, the objective function (5) is restricted by the constraint $\mathbf{U}^T \mathbf{X} \mathbf{X}^T \mathbf{U} = 1$. It is known that Eqn.5 is a constrained quadratic programming problem.

2.2 Multi-view discriminative framework

Based on DNPE, we can obtain discriminative information from facial expression images. However, multi-view facial expression recognition aims to recognize facial expressions with

any views. In this section, we will introduce the multi-view discriminative framework which explores the correlation of facial expressions and views.

In practice, there can be an arbitrary view angle for facial expression recognition. Here we simply suppose that there exist N views for each sample. Given the samples with the i th view ($i \in \{1, \dots, N\}$), they are denoted as \mathbf{X}_i . From Eqn.5, we can obtain the formula $\mathbf{A}_i = (\mathbf{S}_i^{\text{bw}} - \eta_i \mathbf{S}_i^{\text{wi}})$ of the i th view with respect to \mathbf{X}_i . Concerning all views, the optimization problem (i.e. Eqn.5) can be revised as follows,

$$[\mathbf{U}_1, \dots, \mathbf{U}_N] = \arg \max \sum_{i=1}^N \mu_i \mathbf{U}_i^T \mathbf{A}_i \mathbf{U}_i, \quad (6)$$

with constraints $\mathbf{U}_i^T \mathbf{X}_i \mathbf{X}_i^T \mathbf{U}_i = 1, \forall i$, where the positive term μ_i is included to bring the balance between multiple objectives. Because of non-linear constraints, it leads to no closed form solution in the current form. Instead, the relaxed version of the problem can be obtained by coupling all constraints as $\sum_i \mathbf{U}_i^T \mathbf{X}_i \mathbf{X}_i^T \mathbf{U}_i = 1$.

The aforementioned optimization problem embeds the respective objective function of each view into one common function of all views, yet still loses the correlation of samples with distinct views. MCCA [14] was proposed to search the correlation vector for multiple sets. In MCCA, it maximizes all correlations among new variables simultaneously. Therefore, it motivates us to use this method to make the samples with the same expression label yet in various views accumulating very close. According to MCCA theorem and the closeness formulation in [1], we define the objective function that maximizes the correlation between samples with the same expression label yet on all views and meanwhile minimizes the covariance of samples with the same view label, as follows,

$$[\mathbf{U}_1, \dots, \mathbf{U}_N] = \arg \max \sum_{i=1}^N \sum_{j=1, j \neq i}^N \mathbf{U}_i^T \mathbf{M}_i \mathbf{M}_j^T \mathbf{U}_j, \quad (7)$$

with the constraint $\sum_{i=1}^N \mathbf{U}_i^T \mathbf{X}_i \mathbf{X}_i^T \mathbf{U}_i = 1$, where \mathbf{M}_i is the class mean matrix of samples on the i th view.

Based on two objective functions (6) and (7), we can finally obtain the completed formulation as follows,

$$[\mathbf{U}_1, \dots, \mathbf{U}_N] = \arg \max \sum_{i=1}^N \{ \mu_i \mathbf{U}_i^T \mathbf{A}_i \mathbf{U}_i + \sum_{j=1, j \neq i}^N \alpha_{i,j} \mathbf{U}_i^T \mathbf{M}_i \mathbf{M}_j^T \mathbf{U}_j \}, \quad (8)$$

with the constraint $\sum_{i=1}^N \beta_i \mathbf{U}_i^T \mathbf{X}_i \mathbf{X}_i^T \mathbf{U}_i = 1$, where μ_i , $\alpha_{i,j}$ and β_i are balancing parameters which adjust the importance of terms in the objective function and constraint item.

It is observed that Eqn.8 is a standard generalized eigenvalue problem that can be solved using any eigen-solver. Through this formula, we can obtain discriminative feature space of facial expression \mathbf{U}_i in each view.

2.3 Emotion Classification

Our aim is to match two face images with the same or different facial expression label in different views. Partly motivated by cross-view classification [9], we use the cross-view correlation maximization method for matching the test sample \bar{x} , because the lower dimensional

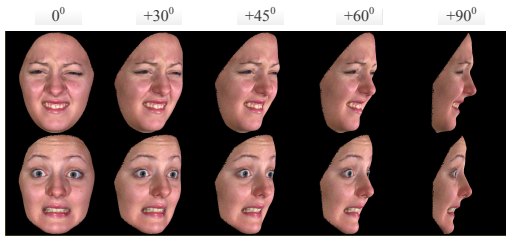


Figure 1: An example of face images from five yaw views (0° , $+30^\circ$, $+45^\circ$, $+60^\circ$, $+90^\circ$) with disgust (top row) and fear (bottom row) expressions from BU-3DFE database [19].

feature space of one specific view has close correlation to other views. Here we design the mean-correlation maximization classifier to classify this sample as follows,

$$h(\vec{x}) = \max_c (\text{mean}(\max_i \{ \text{corr}(\mathbf{U}_i^T \mathbf{X}_{i,c}, \mathbf{U}_i^T \vec{x}) \}_{i=1}^N)), \quad (9)$$

where $\mathbf{X}_{i,c}$ represents training samples of the c th facial expression label with the i th view, corr represents Pearson's linear correlation coefficient operator, N is the number of views, *mean* and *max* are the mean and maximum value operator, respectively.

3 Experimental analysis

In order to evaluate the performance and effectiveness of our method, the BU-3DFE [19] and Multi-PIE [4] databases are used to compare our method with the state of the art.

3.1 BU-3DFE dataset experiments

In the BU-3DFE database, there are 100 subjects with 3D models and face images. By projecting 3D facial expression models in various directions, we can generate a set of 2D facial images with various facial views. In our experiment, we choose 3D models with the highest level of intensity to generate five yaw views (0° , $+30^\circ$, $+45^\circ$, $+60^\circ$, $+90^\circ$) with six facial expressions: anger (An), disgust (Di), fear (Fe), happiness (Ha), sadness (Sa), and surprise (Su). Consequently, we have 3000 face images totally for our experiment. Fig. 1 shows an example of two facial expressions in five yaw views. In experiments, we randomly divided 100 subjects into 10 groups, each one having 10 subjects. In each trial of the experiment, we choose one group as the test set and the other ones as the training set. We conduct ten trials of the experiment in total such that each subject is used as test data once.

We compare our approach with four dimensionality reduction methods (PCA, LDA, LPP and NPE) based on view-dependent framework [7, 8]. Here the local Gabor binary pattern feature operators [21] are used as features in our experiments. In our approach, we feed all training samples of all views to the input of the method. Before feeding it, we use PCA to reduce the data dimension in each view, because the implemented approach could lead to large eigenvalue problems. We keep the top 250 principal components that retained around 90% of the variance. For related parameters in our method, we fix $k^{\text{wi}} = k^{\text{bw}} = 20$, $\alpha_{i,j} = 1000$ for all i and j , $\mu = 1$, and $\beta_i = \frac{\text{tr}(\mathbf{X}_i \mathbf{X}_i^T)}{\text{tr}(\mathbf{X}_i \mathbf{X}_i^T)}$, where \mathbf{X}_i corresponds to training samples at the i th view. More specially, \mathbf{X}_1 represents training samples in the frontal view, when $i = 1$.

Table 1: Performance of various methods on BU-3DFE and Multi-PIE databases (%)

Method	Recognition	
	BU-3DFE	Multi-PIE
PCA	67.70(85)	69.12(80)
LDA	65.87(5)	76.09(4)
LPP	69.33(79)	74.57(20)
NPE	69.63(61)	74.12(15)
DNPE	69.72 (82)	74.35 (30)
Zheng et al. [22]	68.35 (25)	-
BDA/GMM [23]	68.28 (30)	-
Ours	72.47 (35)	76.83 (15)

The first column in Table 1 shows the best results obtained in the optimal feature space and the corresponding dimensionality for each method, where the performance of [22] and [23] is taken directly from their papers. Numbers inside brackets in Table 1 represent the corresponding dimensionalities of the best result. As shown in Table 1, NPE outperforms PCA, LDA, and LPP. DNPE is slightly better than NPE. Our approach outperforms NPE and has best performance among all compared methods. In [22, 23], the same facial expressions from all views were considered as Gaussian mixture, but our method explores the correlation of the same facial expression in different views. So our discriminative space can be obtained more easily by using DNPE and MCCA without Gaussian mixture assumption. Compared to [22, 23], our approach based on DNPE and MCCA provides 4.12% and 4.19% better recognition rate, respectively.

Table 2 shows overall recognition rates as well as the recognition rates of facial expressions of the proposed method across various views. The increasing view angles can affect much on the performance of facial expression recognition. The optimal results for average, anger, disgust, fear, happiness, sadness are achieved when face images are in the frontal view. From Table 2, it can be seen that surprise expression be recognized more easily at angles of 30°, 45° and 60°. These are most likely because the lip movement provides respective evidence to surprise expression in the non-frontal view. It is surprising to see that the performance of sadness at 60° is better than the one at 45°. This is mostly likely due to that the lip movement has more evidence at 60°. We can also see a similar situation for anger at 45° and 60° because of the eye movement.

Table 3 shows more details for six facial expressions with our method. The facial expressions with recognition performance from best to worst are: surprise, happy, anger, disgust, sadness, and fear. Among these expressions, surprise is easiest to be recognized with the recognition rate of 88.80%, while fear is the most difficult to be recognized. This may be caused by that rising eyebrows and dropped jaw are obvious to surprise. However, the tips to fear (e.g. lip slightly stretched, raised upper eyelids and tensor lower eyelids) may be difficult to be captured from face images.

3.2 Multi-PIE dataset experiment

Facial images in the aforementioned experiment are generated from the 3D face model in BU-3DFE. This section is going to build on the previous section by applying our method-

Table 2: Facial expression rate of each view on BU-3DFE database (%)

View	0	30	45	60	90
An	73	72	71	72	67
Di	77	72	68	60	60
Fe	65	58	57	50	49
Ha	91	89	89	86	88
Sa	69	65	60	65	57
Su	89	90	91	90	84
Avg	77.33	74.33	72.67	70.50	67.50

Table 3: Confusion matrix of facial expressions on BU-3DFE database (%).

	An	Di	Fe	Ha	Sa	Su
An	71.00	10.80	3.60	1.80	12.80	0.00
Di	13.00	67.40	5.60	3.00	6.60	4.00
Fe	6.80	8.60	55.80	14.20	11.00	3.60
Ha	0.20	3.20	6.00	88.60	0.40	1.60
Sa	20.00	6.00	8.80	1.20	63.20	0.80
Su	0.60	4.00	3.80	1.60	1.00	88.80

ology on real data. To do this we use the Multi-PIE database to evaluate the performance of the approach presented in this paper. Multi-PIE database has face images from 337 subjects taken across 15 different poses, 20 illuminations, 6 facial expressions (neutral, smile, surprise, squint, disgust and scream) and 4 different sessions. In each session, subjects were instructed to display various facial expressions. In this experiment, we have chosen 100 subjects with 5 facial expressions with the exception of neutral, normal illumination and 13 poses ranging from the left profile (-90°) to the right profile ($+90^\circ$) at an interval of 15° . In order to evaluate our method, all images are cropped and aligned using 4 hand annotated landmark points (eyes, nose tip and mouth), and then normalized to 240 by 240 pixels. Fig. 2 shows examples of post-processed face images from Multi-PIE dataset. In a real application, it can be obtained by face detection [13, 24]. The parameters $\alpha_{i,j}$, β_i , and μ are equal to the experiment in BU-3DFE, and $k^{wi} = k^{bw} = 50$.

The recognition rates of comparative methods in this database are shown in Table 1. From the second column of Table 1, it can be seen that our approach on this database out-

Table 4: Confusion matrix of facial expressions on Multi-PIE dataset (%)

	Disgust	Scream	Smile	Squint	Surprise
Disgust	65.77	1.46	4.00	27.77	1.00
Scream	2.15	83.54	0.69	3.23	10.38
Smile	3.46	0.15	78.69	15.23	2.46
Squint	17.00	1.23	6.54	73.54	1.69
Surprise	1.69	7.54	3.15	5.00	82.62

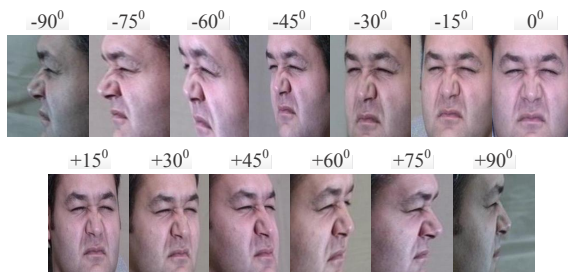


Figure 2: An example of face images (disgust) from Multi-PIE database [4]. Top row—the view variant from -90° to the frontal view. Bottom row—the view variant from $+15^\circ$ to $+90^\circ$.

performs the PCA, LPP and NPE, in which their rates are 7.71%, 2.27% and 2.71% higher, respectively. And our method also performs better than LDA by 0.74%.

Table 4 shows the confusion matrix of 5 facial expressions using our method. The facial expressions with the recognition performance ranked from the best to the worst are following: scream, surprise, smile, squint, disgust. The recognition rate for scream is 83.54%, while for disgust it is 65.77%.

4 Conclusion

Pose variations are challenging to facial expression recognition. In this paper, we propose a multi-view discriminative neighbor preserving embedding approach for recognizing facial expressions in different views. In our method, we propose to exploit the intra-class intrinsic structure and the inter-class penalty graph for strengthening the discriminative power of neighbor preserving embedding. In addition, we present to embed this method into the multi-view model. In order to make intra-class samples with distinct views staying close, MCCA is used to maximize their correlation. These schemes therefore lead to lower dimensional feature spaces having discriminative capability independent to view variations. The proposed method is tested on face images with various views in BU-3DFE and Multi-PIE databases. The experimental results show that our method can achieve a promising recognition performance.

Acknowledgements: This work was sponsored by the Academy of Finland, Infotech Oulu, Nokia Foundation, and Grant 40297/11 from Tekes.

References

- [1] S. Abhishek, K. Abhishek, and D. Hal. Generalized multiview analysis: a discriminative latent space. In *Proc. CVPR*, pages 2160–2167, 2012.
- [2] S. Anvar, W. Yau, and E. Teoh. Multi-view face detection and registration requiring minimal manual intervention. *IEEE Trans. Pattern Anal. Mach. Intell.*, (5), 2013.
- [3] 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.*, 19(7):711–720, 1997.

- [4] R. Gross, I. Matthews, J. Cohn, T. Kanade, and S. Baker. Multi-PIE. *Image Vis. Comput.*, 28(5):807–813, 2010.
- [5] X. He and P. Niyogi. Locality preserving projections. In *Proc. NIPS*, pages 585–591, 2003.
- [6] X. He, D. Cai, S. Yan, and H. Zhang. Neighborhood preserving embedding. In *Proc. ICCV*, pages 1208–1213, 2005.
- [7] Y. Hu, Z. Zeng, L. Yin, X. Wei, J. Tu, and T.S. Huang. A study of non-frontal-view facial expression recognition. In *Proc. ICPR*, pages 1–4, 2008.
- [8] Y. Hu, Z. Zeng, L. Yin, X. Wei, J. Tu, and T.S. Huang. Multi-view facial expression recognition. In *Proc. AFGR*, pages 1–6, 2008.
- [9] A. Li, S. Shan, and W. Gao. Coupled bias-variance trade off for cross pose face recognition. *IEEE Trans. Image Processing*, 21(1):305–315, 2012.
- [10] H. Li, T. Jiang, and K. Zhang. Efficient and robust feature extraction by maximum margin criterion. In *Proc. NIPS*, pages 157–165, 2003.
- [11] D. Liang, J. Yang, Z. Zheng, and Y. Chang. A facial expression recognition system based on supervised locally linear embedding. *Pattern Recognition Letters*, 26(15):2374–2389, 2005.
- [12] S. Liu, Q. Ruan, C. Wang, and G. An. Tensor rank one differential graph preserving analysis for facial expression recognition. *Image Vis. Comput.*, 30(8):535–545, 2012.
- [13] S. Moore and R. Bowden. Local binary patterns for multi-view facial expression recognition. *Computer Vision and Image Understanding*, 115(4):541–558, 2011.
- [14] A. Nielsen. Multiset canonical correlation analysis and multispectral, truly multitemporal remote sensing data. *IEEE Trans. Image Processing*, 11(3):293–305, 2002.
- [15] O. Rudovic, I. Patras, and M. Pantic. Regression-based multi-view facial expression recognition. In *Proc. ICPR*, pages 4121–4124, 2010.
- [16] O. Rudovic, I. Patras, and M. Pantic. Coupled gaussian process regression for pose-invariant facial expression recognition. In *Proc. ECCV*, pages 350–363, 2010.
- [17] C. Shan, S. Gong, and P. McOwan. Appearance manifold of facial expression. In *Proc. ICCV-HCI*, pages 221–230, 2005.
- [18] M.A. Turk and A.P. Pentland. Face recognition using eigenfaces. In *Proc. CVPR*, pages 586–591, 1991.
- [19] L. Yin, X. Wei, Y. Sun, J. Wang, and M. Rosato. A 3d facial expression database for facial behavior research. In *Proc. AFGR*, pages 211–216, 2006.
- [20] Z. Zeng, M. Pantic, G. Roisman, and T.S. Huang. A survey of affective recognition methods: Audio, visual and spontaneous expressions. *IEEE Trans. Pattern Anal. Mach. Intell.*, 27(5):699–714, 2005.

-
- [21] W. Zhang, S. Shan, W. Gao, and X. Chen. Local gabor binary pattern histogram sequence (lgbphs): a novel non-statistical model for face representation and recognition. In *Proc. ICCV*, pages 786–791, 2005.
- [22] W. Zheng, H. Tang, Z. Lin, and T.S. Huang. A novel approach to expression recognition from non-frontal face images. In *Proc. ICCV*, pages 1901–1908, 2009.
- [23] W. Zheng, H. Tang, Z. Lin, and T.S. Huang. Emotion recognition from arbitrary view facial images. In *Proc. ECCV*, pages 490–503, 2010.
- [24] X. Zhu and D. Ramanan. Face detection, pose estimation and landmark localization in the wild. In *Proc. CVPR*, pages 2879–2886, 2012.