

# Misalignment-Robust Face Recognition

Huan Wang<sup>1</sup>      Shuicheng Yan<sup>2</sup>      Thomas Huang<sup>3</sup>      Jianzhuang Liu<sup>1</sup>      Xiaoou Tang<sup>1,4</sup>  
<sup>1</sup> IE, Chinese University of Hong Kong, Hong Kong      <sup>2</sup> ECE, National University of Singapore, Singapore      <sup>3</sup> ECE, University of Illinois at Urbana-Champaign, USA      <sup>4</sup> Microsoft Research Asia Beijing, China  
hwang@ie.cuhk.edu.hk      eleyans@nus.edu.sg      huang@ifp.uiuc.edu      jzliu@ie.cuhk.edu.hk      xitang@microsoft.com

## Abstract

In this paper, we study the problem of subspace-based face recognition under scenarios with spatial misalignments and/or image occlusions. For a given subspace, the embedding of a new datum and the underlying spatial misalignment parameters are simultaneously inferred by solving a constrained  $\ell_1$  norm optimization problem, which minimizes the error between the misalignment-amended image and the image reconstructed from the given subspace along with its principal complementary subspace. A byproduct of this formulation is the capability to detect the underlying image occlusions. Extensive experiments on spatial misalignment estimation, image occlusion detection, and face recognition with spatial misalignments and image occlusions all validate the effectiveness of our proposed general formulation.

## 1. Introduction

Subspace learning techniques for face recognition have experienced a dramatic growth over the past decade. Among them, some popular ones are Principal Component Analysis (PCA) [10], Linear Discriminant Analysis (LDA) [3], Random Subspace [14], Unified Subspace [15], LaplacianFaces [5], Marginal Fisher Analysis [12], Kernel LDA [13], Probabilistic LDA [7], and the recently proposed extensions for handling tensor data [12] [16].

In subspace learning, explicit semantics is assumed for each feature. But for computer vision tasks, *e.g.*, face recognition, the explicit semantics of the features may be degraded by *spatial misalignments*. Face cropping is an inevitable step in an automatic face recognition system, and the success of subspace learning for face recognition relies heavily on the performance of the face detection and face alignment processes. Practical systems, or even manual face cropping, may bring considerable image misalignments, including translations, scaling and rotation, which consequently change the semantics of two pixels with the same index but in different images. Figure 1 demon-

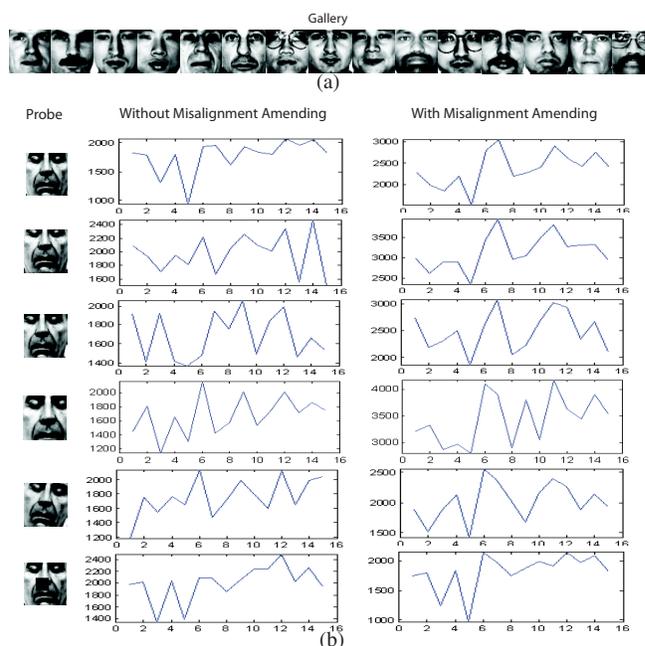


Figure 1. Euclidean distance variations caused by image misalignments: a) example gallery images, b) Euclidean distances between the probe image and the gallery samples indexed from 1 to 15. 1st row: original image, 2nd row: vertical translation, 3rd row: horizontal translation, 4th row: scaling, 5th row: rotation, and 6th row: occlusion. The statistics are computed within the LDA subspace of the YALE database, and the right column is obtained from our proposed misalignment robust algorithm, which effectively overcomes the influence of spatial misalignments and image occlusions.

strates that these spatial misalignments and image occlusions may greatly affect image similarity measurement, and consequently degrade classification performance. Hence it is desirable to have a general solution for misalignment-robust face recognition that is applicable for all the above-mentioned subspace learning algorithms.

In the literature, there exist some attempts to tackle this issue, *e.g.*, in [8], the effect of spatial misalignments was alleviated to some extent by adding virtual training samples

with manual spatial misalignments. But the spatial misalignment issue is still far from being solved, since 1) in the training stage, usually all samples have been cropped out, thus it is impossible to synthesize virtual samples without the original larger-size samples; 2) the added virtual samples will make the data much more inseparable; and 3) the number of virtual samples is limited compared with the huge amount of possible spatial misalignments.

In this paper, we propose a principled solution to the face recognition problem under scenarios with spatial misalignments and/or image occlusions. A unified constrained  $\ell_1$  norm optimization formulation, generally applicable for any learnt subspace, is proposed to infer the embedding of a new datum in the learnt subspace and at the same time estimate the spatial misalignment parameters as well as the possible image occlusions. Consequently we achieve algorithmic robustness to spatial misalignment and image occlusion for face recognition. The constraints of the  $\ell_1$  norm optimization problem impose the feasibility of obtaining the misalignment parameters. The objective function measures the difference between the misalignment-amended image and the image reconstructed from the learnt subspace as well as its principal complementary subspace. The minimization of the  $\ell_1$  norm of this difference ensures that the border areas and the possibly occluded area of the new datum will have less effect on the estimation of the parameters for the subspace and spatial misalignments.

## 2. Background and Motivation

Face recognition, as a classic multi-class pattern recognition problem, has been very popular for validating the effectiveness of newly proposed subspace learning algorithms and classification approaches. In this section, we first give a brief overview of subspace learning, and then introduce the spatial misalignment issue specifically suffered by visual classification tasks.

### 2.1. Subspace Learning Overview

For face recognition, let the training data be  $\{x_i | x_i \in \mathbb{R}^m\}_{i=1}^N$ , where  $N$  is the number of training samples and the data are assumed to be zero centered. The corresponding subject indices of the samples are denoted as  $\{c_i | c_i \in \{1, 2, \dots, N_c\}\}_{i=1}^N$ , where  $N_c$  is the number of subjects. In practice, dimensionality reduction is in great demand owing to the fact that the effective information for classification often lies within a much lower dimensional feature space.

A simple but effective approach to dimensionality reduction is to find a matrix  $W = [w_1, w_2, \dots, w_d] \in \mathbb{R}^{m \times d}$  ( $\text{Rank}(W) = d$ ,  $\|w_k\| = 1$ ,  $k = 1, 2, \dots, d$ ) to transform the original high-dimensional data  $x$  into a low-dimensional form  $y \in \mathbb{R}^d$  (usually  $d \ll m$ ) as

$$y = W^T x, \quad (1)$$

where the column vectors of the matrix  $W$  constitute a subspace for data representation. Subspace learning algorithms are designed to search for such a matrix.

### 2.2. Motivation

Assume that a projection  $W$  has been derived from a certain subspace learning algorithm. When a new datum  $x$  comes, generally it is directly projected into the learnt subspace spanned by the column vectors of  $W$  as in (1). However, for computer vision tasks, *e.g.*, face recognition, the face image needs first to be cropped out from the original whole image which possibly contains background. A naive way to perform this is to fix the locations of the two eyes in the cropped rectangle [12]. But for practical systems, the positions of the two eyes need be automatically located by a face alignment algorithm [4] or eye detector [11], so it is inevitable that there may exist localization errors, namely spatial misalignments. Generally, the spatial misalignments include four components, translations in horizontal and vertical directions  $(T_x, T_y)$ , scaling  $(r)$ , and rotation  $(\alpha)$ . Mathematically, the underlying face image  $\hat{x}$  without spatial misalignments can be considered as the transformed face image by a matrix  $P$  from the cropped face image  $x$ , and then the exact low-dimensional representation is

$$W^T \hat{x} = W^T P x, \quad (2)$$

which is not exactly the same as  $W^T x$ . Their difference is

$$\hat{\varepsilon} = W^T \hat{x} - W^T x = W^T (P - I)x. \quad (3)$$

Here, an empirical evaluation of the effect from  $\hat{\varepsilon}$  to the data metric measurement is presented in Figure 1. We can see that the spatial misalignments may greatly affect the metric measurement within the learned subspace. This motivates the need for a general procedure to infer the representation of a new datum within a certain learnt subspace in a way robust to spatial misalignments.

## 3. Misalignment Robust Face Recognition

In this section, we present our solution to misalignment-robust face recognition. More specifically speaking, when a new datum comes, its embedding in the subspace spanned by  $W$  and the underlying image misalignment parameters are simultaneously inferred, and consequently the datum is essentially projected from the misalignment-amended image.

### 3.1. Problem Formulation

**Image reconstruction from  $W$  and its principal complementary subspace.** Let  $x$  be a new datum, which may contain image misalignments. We use the generative model

to estimate the parameters describing the spatial misalignments. As the matrix  $W$  may be learnt for various purposes, such as discriminating power [3][12] and locality preservation [5], it may be unnecessary to be best at generating the original datum. To reconstruct the underlying misalignment-amended image of  $x$ , we introduce another subspace spanned by  $W^\sharp \in \mathbb{R}^r$ , called the principal complementary subspace of  $W$ .

The matrix  $W^\sharp$  is learnt as follows. First, we remove the information covered by the matrix  $W$  for all the training data as

$$x_i^r = x_i - W^\dagger W^T x_i, \quad (4)$$

where  $W^\dagger$  is the pseudo-inverse of the matrix  $W$  and used to transform the low-dimensional representation back to the original feature space. Note that the training data are assumed to be zero centered, and hence the above equation does not include the data mean term. Then, the column vectors of  $W^\sharp$  are computed as the principal components of the covariance matrix  $C^r$  from the remainder data  $x_i^r$ 's, where

$$C^r = \frac{1}{N} \sum_{i=1}^N x_i^r x_i^{rT}. \quad (5)$$

Finally, the misalignment-amended version  $\hat{x}$  of the datum  $x$  is set to be reconstructed from these two subspaces as

$$\hat{x} = [W, W^\sharp] \begin{bmatrix} y \\ y^\sharp \end{bmatrix} + \varepsilon, \quad (6)$$

where  $y \in \mathbb{R}^d$  and  $y^\sharp \in \mathbb{R}^r$  are the coefficient vectors for the two basis matrices  $W$  and  $W^\sharp$ , and  $\varepsilon$  represents noise. Our task is to infer the vector  $y$  and then use it for final face recognition.

*Discussion:* Although PCA is theoretically optimal in data reconstruction, we do not directly use PCA in this work because the column vectors of  $W$  may not lie within the subspace spanned by PCA, and the reconstructed image from PCA then loses the information useful for the specific purpose characterized by the learnt  $W$ .

**Misalignment-amended image.** As mentioned above, the underlying misalignment-amended image of  $x$  can be considered as the image transformed by matrix  $P$  from the observed image  $x$ . In this work, we do not explicitly use the four parameters  $\theta=(T_x, T_y, r, \alpha)$  to model the spatial misalignments. Instead we simplify this model to assume that each pixel within the misalignment-amended image is the nonnegative linear combination of its neighboring pixels within the observed image  $x$ . More specifically, we assume that the misalignment only affects a  $k_s$ -by- $k_s$  local neighborhood for each pixel. We divide the face image plane into  $n$  blocks of size  $k$ -by- $k$  with  $m = n \times k^2$ , and assume that the same linear combination coefficients apply to all the pixels within each block. We arrange the elements of the

image vector  $x$  block by block, and then the misalignment-amending process can be defined as

$$T_\theta(x) = \text{diag}\{(P_\theta \otimes e_{k^2})N_x\}, \quad (7)$$

where  $P_\theta \in \mathbb{R}^{n \times k_s^2}$  and each row of  $P_\theta$  represents a set of linear combination coefficients for a block;  $e_{k^2}$  is a  $k^2$  dimensional column vector with all ones;  $\otimes$  is the Kronecker Product, defined as  $A \otimes B = [A_{ij}B]$  where  $A = \{A_{ij}\}$  and  $B$  are two arbitrary matrices;  $N_x \in \mathbb{R}^{k_s^2 \times m}$ , with each column vector representing the gray level values (in image  $x$ ) of the  $k_s^2$  nearest neighbors of a pixel; and  $\text{diag}\{\cdot\}$  denotes a vector consisting of the diagonal elements of a square matrix. Then, we have the misalignment-amended image  $\hat{x} = T_\theta(x)$ , i.e.,

$$[W, W^\sharp] \begin{bmatrix} y \\ y^\sharp \end{bmatrix} + \varepsilon = \text{diag}\{(P_\theta \otimes e_{k^2})N_x\}. \quad (8)$$

**Parameter estimation from the  $\ell_1$  norm minimization formulation.** In (8), there exist three sets of parameters to estimate, namely, subspace coefficients ( $y$  and  $y^\sharp$ ), noise vector ( $\varepsilon$ ), and spatial misalignment parameters ( $P_\theta$ ). Eqn. (8) itself is insufficient for inferring the solution of the subspace and the spatial misalignment parameters, and the scaling of the solution will again be the solution of (8). To derive a feasible and reasonable solution, on the one hand, the misalignment parameters should be non-negative, that is,

$$P_\theta \geq 0, \quad (9)$$

and the linear combination coefficients for a certain pixel should sum up to one, namely,

$$P_\theta e_{k_s^2} = e_n, \quad (10)$$

where  $e_{k_s^2}$  and  $e_n$  are  $k_s^2$  and  $n$  dimensional column vectors respectively with all ones.

On the other hand, when the neighboring pixels are out of the image plane for a certain pixel, we generally use zero values to fill in these areas, and consequently, there may exist very large errors within  $\varepsilon$  for the pixels near the boundary. Moreover, the image occlusions and noises may also result in large values for the elements of  $\varepsilon$ .

A natural way to obtain a solution robust to the above factors is to minimize the  $\ell_1$  norm of the error term  $\varepsilon$ , such that the large errors only appear on the pixels near the boundary or with possible occlusions/noise.

To sum up the above objective function and all constraints, we have the formulation for the subspace and misalignment parameter estimation as listed in Algorithm 1. It is a general  $\ell_1$  norm optimization problem with variables ( $y, y^\sharp, \varepsilon, P_\theta$ ). This problem is convex and can be transformed into a general linear programming problem by adding extra auxiliary coefficients. Hence there exists a

---

**Algorithm 1** Procedure for simultaneously inferring subspace and misalignment parameters

---

Minimize:  $\|\varepsilon\|_1$ , *s.t.*

- 1:  $[W, W^\sharp] \begin{bmatrix} y \\ y^\sharp \end{bmatrix} + \varepsilon = \text{diag}\{(P_\theta \otimes e_{k^2})N_x\}$ ;
  - 2:  $P_\theta e_{k^2} = e_n$ ;
  - 3:  $P_\theta \geq 0$ .
- 

globally optimal solution. In practice we drop the last non-negative constraint and incorporate the second constraint into the objective with a large penalty coefficient. Then the optimization is solved efficiently using the general linear programming toolbox or  $\ell_1$  norm optimization toolbox as in [1].

**Byproduct: Occlusion Detection.** A byproduct of this formulation is that, when there exist image occlusions in the observed face image  $x$ , the  $\ell_1$  norm minimization of the error vector  $\varepsilon$  can also recover these areas as the pixels with large errors in  $\varepsilon$ . For the case with image occlusions, after we detect the occlusion area, the subspace parameters  $y$  and  $y^\sharp$  can be further refined by replacing the occluded pixels with the values from the reconstructed image based on the subspaces spanned by  $W$  and  $W^\sharp$ .

### 3.2. Discussions

In this subsection, we discuss the relationship between our proposed general formulation for misalignment-robust face recognition and two related works [8][9].

#### 3.2.1 Relationship with [8] using virtual samples

Shan *et al.* [8] proposed to add virtual training samples with manual misalignments for bridging the distribution gap between training data without spatial misalignments and testing data with spatial misalignments. Our formulation in this work is different from [8] in several aspects: 1) the work [8] cannot handle image occlusion; 2) the work [8] cannot work under scenarios where the training images are already cropped; 3) the virtual samples essentially make the classification hyperplane more nonlinear and thus maybe beyond the capability of linear subspace techniques; 4) it cannot estimate the exact spatial misalignment parameters or occluded areas; and 5) our proposed formulation is general and can be used under scenarios with both spatial misalignments and image occlusions, and it can also be used for both misalignment and occlusion estimation. Moreover, our formulation can also work on the derived subspace from the training set with virtual samples to further improve algorithmic performance on testing data with unforeseen spatial misalignments.

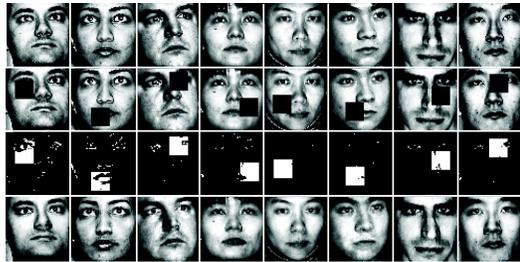


Figure 2. Demonstration of occlusion detection on the CMU PIE database. Original samples are displayed in the first row. An 18-by-18 occlusion is randomly generated as shown in the second row. The third row shows the error maps derived from our algorithm, and the recovered images are demonstrated in the bottom row.

#### 3.2.2 Relationship with the shift invariant PCA [9]

Tu *et al.* [9] proposed a shift invariant probabilistic PCA for alleviating the influence of image shifts, namely spatial translations, on PCA based face recognition. This algorithm is specific to PCA and limited in the following aspects compared with our formulation: 1) the work can only handle image translations, while cannot handle other types of spatial misalignments; 2) the work is specific to generative algorithms, and cannot be used for discriminative algorithms, such as LDA and MFA; 3) similar to [8], it cannot handle the cases with image occlusions. Our experiments are designed for supervised subspace learning, which generally can provide better results than unsupervised algorithms.

Since the methods in [8] and [9] are too limited to handle the cases in our experiments, we do not compare our formulation with them in the experimental section.

## 4. Experiments

In this section, we systematically evaluate the effectiveness of our general formulation for misalignment-robust (MAR) face recognition, and here we take two popular subspace learning algorithms, LDA [3] and MFA [12], as examples for the evaluation. The evaluation consists of three aspects: 1) occlusion detection and recovery, 2) face recognition on testing data with synthesized spatial misalignments or image occlusions, and 3) face recognition under the scenario with automatic image cropping.

### 4.1. Data Sets

Four benchmark face databases, *i.e.*, ORL, CMU PIE, YALE databases<sup>1</sup>, and the Face Recognition Grand Challenge database (FRGC version 1.0) [2] are used in our experiments. The ORL database contains 400 images of 40 persons, where each image is manually cropped and normalized to the size of 32-by-28 pixels. The CMU PIE (Pose,

<sup>1</sup>Available at <http://www.face-rec.org/databases/>.

Table 1. Recognition accuracy rates (%) on the three databases: manually aligned images vs. images with mixed misalignments.

Configuration	Baselines		LDA Related Algorithms		MFA Related Algorithms	
<b>YALE</b>	w/o DR	PCA	Ori-LDA	MAR-LDA	Ori-MFA	MAR-MFA
N6T5	78.7/56.0	82.7/60.0	89.3/68.0	<b>94.6/78.7</b>	90.7/68.0	93.3/ <b>81.3</b>
N5T6	72.2/52.2	72.2/53.3	82.2/63.3	90.0/ <b>73.3</b>	82.2/62.2	<b>92.2/72.2</b>
N4T7	72.4/54.3	72.4/53.3	82.9/61.0	<b>88.6/75.2</b>	83.8/61.9	<b>88.6/73.3</b>
<b>ORL</b>	w/o DR	PCA	Ori-LDA	MAR-LDA	Ori-MFA	MAR-MFA
N4T6	87.9/64.2	88.0/63.2	88.3/51.7	89.6/65.7	89.2/51.2	<b>90.0/69.6</b>
N3T7	81.4/52.9	81.8/53.9	84.3/50.4	85.0/ <b>65.7</b>	83.6/48.9	<b>86.1/64.6</b>
N2T8	71.6/46.9	68.8/49.1	71.3/45.3	75.6/54.4	72.2/45.9	<b>76.3/55.0</b>
<b>PIE</b>	w/o DR	PCA	Ori-LDA	MAR-LDA	Ori-MFA	MAR-MFA
N4T6	84.4/62.2	87.8/65.9	92.9/54.0	94.2/78.8	93.9/55.0	<b>94.4/79.6</b>
N3T7	80.7/54.2	83.5/55.9	94.1/50.3	94.1/72.6	<b>95.0/51.5</b>	<b>95.0/73.7</b>
N2T8	78.8/51.8	81.8/51.6	84.7/46.2	89.1/69.6	86.7/46.4	<b>89.9/70.8</b>

Illumination, and Expression) database contains more than 40,000 facial images of 68 people. In our experiment, a subset of five near frontal poses (C27, C05, C29, C09 and C07) and illuminations indexed as 08 and 11 are used and manually normalized to the size of 32-by-32 for the face recognition experiments. The Yale face database contains 165 grayscale images of 15 individuals with 11 images per subject, one per different facial expression or configuration: center-light, with/without glasses, happy, left-light, normal, right-light, sad, sleepy, surprised, and wink. The images are also manually cropped and normalized to the size of 32-by-32 pixels. The FRGC database consists 5658 images of 275 subjects. The number of facial images of each subject varies from 6 to 48. For the FRGC database, we randomly select half of the images of each person for model training, and the left half for testing. No information for manual cropping is available for FRGC database, and hence instead the images are automatically cropped and then normalized to the size of 32-by-32 pixels in the experiments.

## 4.2. Occlusion Detection

For facial images with occlusions, the occluded parts can be revealed by detecting the elements of  $\varepsilon$  with relatively large reconstruction errors. In this subsection we examine the occlusion detection capability of our MAR formulation on the CMU PIE database. We randomly pick 4 images of each subject for training the subspace to derive  $W$  and  $W^\sharp$ . The remaining 6 images of each person serve as probe images. Similar to the spatial misalignment estimation experiments, we normalize the images to a larger size of 64-by-64 pixels and then an 18-by-18 artificial occlusion is generated at a random position. Correspondingly, we select  $18 \times 18 = 324$  pixels with the largest  $\varepsilon$ 's as the occluded pixels. Eight images are randomly selected from the probe set and the occlusion detection results are shown in Figure 2, from which we observe that the positions of the occluded parts are generally recognized. Consequently,

the facial images without occlusions can be further reconstructed from Eqn.(6), which is demonstrated in Figure 2. The configuration for the subspace learning algorithm is the same as that for the spatial misalignment estimation.

## 4.3. Face Recognition with Misalignments

In this subsection, face recognition experiments are conducted on three benchmark face databases with spatial misalignments for the testing data. Our MAR framework is evaluated based on two popular subspace learning algorithms, LDA and MFA. For the MFA related algorithms, the number of intra-class nearest neighbors of each sample is fixed as 3, and the number of closest inter-class pairs for each class is set to 40 for CMU PIE and ORL. For the Yale database, the latter number is set to 10 since the class number is comparably smaller for this database. To speed up model training and avoid the singularity problem, PCA is conducted as a preprocessing step for the original LDA and MFA. Similar to the Fisherface algorithm [3], the PCA dimension is set to  $N - N_c$ , where  $N$  is the sample number and  $N_c$  is the class number.

For comparison, the classification results on the original gray-level features without dimensionality reduction are also reported as the baseline, denoted as 'w/o DR' in the result tables. In all the experiments, the Nearest Neighbor method is used for final classification. All possible dimensions of the final low-dimensional representation are evaluated, and the best results are reported. For each database, we test various configurations of training and testing sets for the sake of statistical confidence, denoted as ' $N_x T_y$ ' for which  $x$  images of each subject are randomly selected for model training and the remaining  $y$  images of each subject are used for testing. We used the mixed spatial misalignments to simulate the misalignments brought by the automatic face alignment process. In the mixed spatial misalignment configuration, a rotation  $\alpha \in [-5^\circ, +5^\circ]$ , a scaling  $r \in [0.95, 1.05]$ , a horizontal shift  $T_x \in [-1, +1]$  and

Table 2. Recognition accuracy rates (%) on automatically cropped images (for both training and testing data).

Configuration	Baselines		LDA Related Algorithms		MFA Related Algorithms	
	w/o DR	PCA	Ori-LDA	MAR-LDA	Ori-MFA	MAR-MFA
<b>FRGC</b>						
50%:50%	57.4	57.5	86.0	<b>90.0</b>	86.0	89.5
<b>YALE</b>						
N6T5	80.0	78.7	<b>89.3</b>	<b>89.3</b>	<b>85.3</b>	<b>85.3</b>
N5T6	72.2	68.9	78.9	<b>82.2</b>	78.9	<b>82.2</b>
N4T7	74.3	70.5	79.1	<b>84.8</b>	80.0	83.8

a vertical shift  $T_y \in [-1, +1]$  are randomly added to the original image. The detailed results with mixed spatial misalignments is demonstrated in Table 1 compared with the performance on manually cropped images. From the table, we can have the observations: 1) for all the experiments, the MAR framework combined with LDA and MFA both greatly improve the face recognition accuracy; 2) the MFA based algorithms generally outperform the LDA based algorithms, the performance of which greatly relies on the assumption of the Gaussian distribution for the data; 3) the unsupervised learning algorithm PCA is more robust to spatial misalignments than the supervised algorithms LDA and MFA, especially on the ORL and PIE databases; and 4) under the manual cropping scenario, the recognition rate on the YALE gets a dramatic increase from the MAR framework, while for the other two databases, the improvement is not so obvious. An explanation of which is that the spatial misalignments exist in the YALE database even though they are manually cropped.

#### 4.4. Scenario with Automatic Cropping

Finally, we examine the performance of the MAR framework under the scenario with automatic cropping. We utilized the Active Appearance Model [4] as the face alignment algorithm for automatically locating the key points on the face, and then cropped the face based on the detected key points. The face alignment was conducted on the YALE and FRGC databases for automatic cropping, while for the PIE and ORL databases, automatic face alignment is unavailable since the faces have been cropped out for the ORL database and the face alignment results are unacceptable on the PIE database due to the influence of illuminations.

Detailed results are listed in Table 2, from which the observations can be made: 1) the recognition accuracies under the scenario with automatic cropping are decreased for almost all the algorithms on the YALE database compare with the scenario with manual cropping; and 2) the MAR framework can greatly compensate the effect of spatial misalignments caused by the automatic cropping process.

## 5. Conclusions and Future Work

In this paper, a general  $\ell_1$  norm minimization formulation has been proposed for misalignment-robust face recog-

nition based on subspace learning techniques. In this formulation, the embedding of a new datum in the learnt subspace and the spatial misalignment parameters are simultaneously estimated, and the image occlusion areas may also be detected based on the  $\ell_1$  norm minimization of the difference between the misalignment-amended image and the reconstructed image from the learnt subspace along with its principal complementary subspace. In the future, we plan to take pose variation as a type of specific spatial misalignment, and propose a more general formulation to handle face recognition with pose variations.

## 6. Acknowledgement

The work described in this paper was supported in part by the U.S. Government VACE program and in part by a grant from the Research Grants Council of the Hong Kong SAR, China (Project No. CUHK 414306).

## References

- [1] <http://www.acm.caltech.edu/11magic/> 4
- [2] P. J. Phillips, P. J. Flynn, T. Scruggs, K. W. Bowyer, J. Chang, K. Hoffman, J. Marques, J. Min, and W. Worek. Overview of the face recognition grand challenge. *CVPR*, 2005. 4
- [3] P. Belhumeur, J. Hespanha, and D. Kriegman. Eigenfaces vs. fisherfaces: Recognition using class specific linear projection. *PAMI*, vol. 19, no. 7, pp. 711–720, 1997. 1, 3, 4, 5
- [4] T. Cootes, G. Edwards, and C. Taylor. Active appearance models. *PAMI*, vol. 23, no. 6, pp. 681–685, 2001. 2, 6
- [5] X. He, S. Yan, Y. Hu, P. Niyogi, and H. Zhang. Face Recognition Using Laplacianfaces. *PAMI*, vol. 27, no. 3, 2005. 1, 3
- [6] I. Joliffe. Principal Component Analysis, *Springer-Verlag*, 1986.
- [7] S. Prince and J. Elder. Probabilistic Linear Discriminant Analysis for Inferences About Identity. *CVPR*, 2007. 1
- [8] S. Shan, Y. Chang, W. Gao, B. Cao, and P. Yang. Curse of Mis-alignment in Face Recognition: Problem and a Novel Mis-alignment Learning Solution. *AFGR*, pp. 314–320, 2004. 1, 4
- [9] J. Tu, A. Ivanovic, X. Xu, L. Fei-Fei, and T. Huang. Variational Shift Invariant Probabilistic PCA for Face Recognition. *ICPR*, pp. 548–551, 2006. 4
- [10] M. Turk and A. Pentland. Face recognition using eigenfaces. *CVPR*, pp. 586–591, 1991. 1
- [11] P. Wang, M. Green, Q. Ji, and J. Wayman. Automatic Eye Detection and Its Validation. *CVPR*, vol. 3, pp. 164–171, 2005. 2
- [12] S. Yan, D. Xu, B. Zhang, H. Zhang, Q. Yang and S. Lin. Graph Embedding and Extensions: A General Framework for Dimensionality Reduction. *PAMI*, vol. 29, no. 1, pp. 40–51, 2007. 1, 2, 3, 4
- [13] J. Yang, A. Frangi, J. Yang, D. Zhang, and Z. Jin. KPCA Plus LDA: A Complete Kernel Fisher Discriminant Framework for Feature Extraction and Recognition. *PAMI*, vol. 27, no. 2, pp. 230–244, 2005. 1
- [14] X. Wang and X. Tang. Random Sampling for Subspace Face Recognition. *IJCV*, vol. 79, no. 1, pp. 91–104, 2006. 1
- [15] X. Wang and X. Tang. A unified framework for Subspace Face Recognition. *PAMI*, vol. 26, no. 9, pp. 1222–1228, 2004. 1
- [16] J. Ye, R. Janardan, and Q. Li. Two-Dimensional Linear Discriminant Analysis. *NIPS*, 2004. 1