Misalignment-Robust Face Recognition

Shuicheng Yan[1], Huan Wang[2], Jianzhuang Liu[2], Xiaoou Tang[2], Thomas S. Huang[3]

[1] Department of Electrical and Computer Engineering, National University of Singapore, Singapore
[2] Department of Information Engineering, The Chinese University of Hong Kong, Hong Kong
[3] Department of Electrical and Computer Engineering, University of Illinois at Urbana-Champaign, USA

***Abstract*— Subspace learning techniques for face recognition have been widely studied in the past three decades. In this paper, we study the problem of general subspace-based face recognition under the scenarios with spatial misalignments and/or image occlusions. For a given subspace derived from training data in a supervised, unsupervised, or semi-supervised manner, the embedding of a new datum and its underlying spatial misalignment parameters are simultaneously inferred by solving a constrained ℓ_1 norm optimization problem, which minimizes the ℓ_1 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/or image occlusions all validate the effectiveness of our proposed general formulation for misalignment-robust face recognition.**

I. Introduction

Subspace learning techniques for face recognition have experienced a dramatic growth over the past decade [5] [7] [23] [25]. Among them, some popular ones are Principal Component Analysis (PCA) [16], Linear Discriminant Analysis (LDA) [3], Random Subspace [18], Unified Subspace [19], LaplacianFaces [8], Marginal Fisher Analysis [21], Kernel LDA [23], Probabilistic LDA [11], and the recently proposed extensions for handling tensor data [21] [24]. Subspace learning was originally motivated for overcoming the curse of dimensionality in the learning process and reducing the computational cost for practical applications. Then subspace learning was further proved to be possible and necessary from the fact that the data in a certain application often lie on or nearly on a lower-dimensional manifold. Recently, beyond the different motivations of these popular subspace learning algorithms, most of them were claimed to be unified within a general framework called graph embedding [21].

Subspace learning is a powerful tool widely used in a variety of research areas. Generally 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 demonstrates 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 to all the above-mentioned subspace learning algorithms.

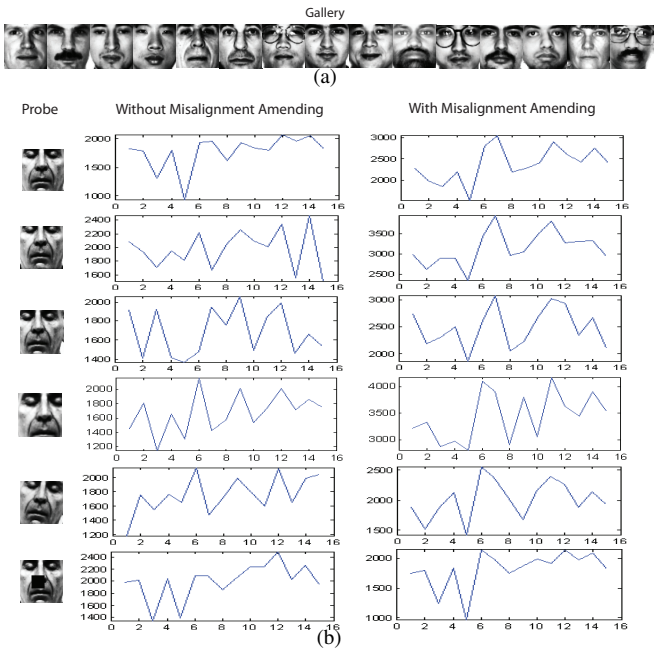

Fig. 1. Euclidean distance variations caused by image misalignments. a) Example gallery images. b) Euclidean distances between the probe image and the gallery samples indexed from 1 to 15, 1st row: original image, 2nd row: vertical translation, 3rd row: horizontal translation, 4th row: scaling, 5th row: 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 the spatial misalignments and image occlusions.

In the literature, there exist some attempts to analyze [13] and tackle this problem, *e.g.*, in [12], the effect of spatial misalignments was alleviated to some extent by adding virtual training samples with manual spatial misalignments. However, the spatial misalignment problem is still far from being solved, since 1) in the training stage, usually all samples have been cropped out, and the virtual sample synthesis process may bring noises to the pixels near image borders; 2) the added virtual samples may make the data more inseparable; and 3) the number of virtual samples is limited compared with the huge amount of possible spatial misalignments. The work

in [10] instead used patch-based philosophy for overcoming misalignment issue.

In this paper, we provide our solution to the face recognition problem under the scenarios with spatial misalignments and/or image occlusions. A unified constrained ℓ_1 norm optimization formulation, generally applicable to 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 ℓ_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 ℓ_1 norm of this difference ensures that the border areas and the possibly occluded area of the new datum have less effect on the estimation of the parameters for the subspace and spatial misalignments.

The rest of this paper is organized as follows. Section II introduces the related work and the motivation of this work. The details of the ℓ_1 norm minimization formulation for misalignment-robust face recognition are described in Section III. Section IV presents extensive comparison experiments on three benchmark face databases, and concluding remarks are given in Section V.

II. 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.

A. 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 $\{c_i|c_i \in \{1,2,...,N_c\}\}_{i=1}^N$, 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 $W = [w_1, w_2, ..., w_d] \in \mathbb{R}^{m\times d}$ (Rank(W) = d, $\|w_k\|$ = 1, k = 1, 2, ..., d) to transform the original high-dimensional data x into a low-dimensional form $y \in \mathbb{R}^d$ (usually $d \ll m$) as

$$y = W^T x, \tag{1}$$

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.

Principal Component Analysis (PCA) [9] seeks projection directions with maximal variances, namely with the best capability to reconstruct the original data. LDA [3] [14] and its variants [26] [22] search for the directions that are most effective for discrimination by minimizing the ratio between the intra-class and inter-class scatters. Locality Preserving Projection (LPP) [8] tries to preserve the local neighborhood relations across the data manifold through a projection matrix W. Marginal Fisher Analysis (MFA) [21], derived from the graph embedding framework, maximizes the ratio of projected distances of between-class marginal points to that of within-class neighboring points. According to the graph embedding framework introduced in [21], most state-of-the-art algorithms for subspace learning can be unified as an optimization of the ratio:

$$\arg\min_W \frac{\sum_{i\neq j} \|W^T x_i - W^T x_j\|^2 S_{ij}}{\sum_{i\neq j} \|W^T x_i - W^T x_j\|^2 S^p_{ij}}, \tag{2}$$

where the weight matrix $S = [S_{ij}]$ describes the relationships between sample pairs that we try to preserve in subspace learning, while $S^p = [S^p_{ij}]$ characterizes the unfavorable relationships that should be avoided.

B. 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 [21]. For practical systems, however, the positions of the two eyes need be automatically located by a face alignment algorithm [6] or eye detector [17], 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 (α). 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

$$W^T \hat{x} = W^T P x, \tag{3}$$

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

$$\hat{\varepsilon} = W^T \hat{x} - W^T x = W^T (P - I) x. \tag{4}$$

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.

III. 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.

A. Problem Formulation

Image reconstruction from W and its principal complementary subspace. Let x be a new datum, which may contain image misalignments. We use a generative model to estimate the parameters describing the spatial misalignments. As the matrix W may be learnt for varying purposes, such as discriminating power [3][21] and locality preservation [8], it is unnecessary to be best, or even possible to be not good, at reconstructing the original datum. Thus, we introduce another subspace spanned by $W^\sharp \in \mathbb{R}^{m\times r}$, called the principal complementary subspace of W, to reconstruct the underlying misalignment-amended image of x along with the learned matrix 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

$$x_i^r = x_i - W^\dagger W^T x_i, \tag{5}$$

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

$$C^r = \frac{1}{N}\sum_{i=1}^{N} x_i^r x_i^{rT}. \tag{6}$$

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

$$\hat{x} = [W,\ W^\sharp]\left[\begin{array}{c} y \\ y^\sharp \end{array}\right] + \varepsilon, \tag{7}$$

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 ε 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 the principal components, 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 θ=(T_x, T_y, r, α) 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

$$T_\theta(x) = diag\{(P_\theta \otimes e_{k^2})N_x\}, \tag{8}$$

where $P_\theta \in \mathbb{R}^{n\times k_s^2}$ and each row of P_θ 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 $diag\{\cdot\}$ 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.*,

$$[W,\ W^\sharp]\left[\begin{array}{c} y \\ y^\sharp \end{array}\right] + \varepsilon = diag\{(P_\theta \otimes e_{k^2})N_x\}. \tag{9}$$

For the better understanding of the relationships between different variables, the example images for these variables are shown in Figure 2.

Parameter estimation from the ℓ_1 norm minimization formulation. In (9), there exist three sets of parameters to estimate, namely, subspace coefficients (y and $y^\sharp$), noise vector (ε), and spatial misalignment parameters (P_θ). Eqn. (9) itself is insufficient for inferring the solution of the subspace and the spatial misalignment parameters, and the scaling of the solution is also the solution of (9). To derive a feasible and reasonable solution, on the one hand, the misalignment parameters should be non-negative, that is,

$$P_\theta \geq 0, \tag{10}$$

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

$$P_\theta\ e_{k_s^2} = e_n, \tag{11}$$

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

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 ε for the pixels near the boundary. Moreover, the image occlusions and noises may also result in large values for the elements of ε. In these scenarios, the number of pixels with large-value noises is relatively small compared with the total number of pixels, namely, the ε is sparse. As studied in [4] and [20], a sparse solution can be achieved by minimizing the ℓ_1 norm. Thus a natural way to obtain a solution robust to the above factors is to minimize the ℓ_1 norm of the error term ε, such that the large errors only appear on the pixels near the boundary or with possible occlusions/noise [20].

Fig. 2. Example W, $W^\sharp$, x (with misalignments mainly in horizontal direction), $\hat{x}$, and ε based on images from the CMU PIE database used in the experiment part, and the subspace is obtained from Linear Discriminant Analysis.

Algorithm 1 Procedure for simultaneously inferring the subspace and misalignment parameters

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

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

2: $P_\theta\ e_{k_s^2} = e_n$;

3: $P_\theta \geq 0$.

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 ℓ_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 globally optimal solution. In practice we drop the last nonnegative constraint and incorporate the second constraint into the objective with a large penalty coefficient. Then the optimization can be solved efficiently using the general linear programming toolbox or ℓ_1 norm optimization toolbox as in [1].

B. Misalignment Estimation

After obtaining the low-dimensional feature representation y corresponding to W for each sample x, face recognition can then be conducted based on simple yet effective Nearest Neighbor approach or other more complicated classifiers.

Besides face recognition, the estimation of image spatial misalignments is also very useful. From the formulation in Algorithm 1, we can obtain the misalignment parameter matrix P_θ. This matrix characterizes the image misalignments, including translations (T_x, T_y), rotation (α), and scaling (r), however the θ=(T_x, T_y, r, α) cannot be directly inferred from P_θ due to the simplification aforementioned. Instead, we first obtain a set of pixel pairs (original pixel vs. misalignment-amended pixel) based on P_θ, and then use these pixel pairs to infer the θ=(T_x, T_y, r, α). More specifically, for the ith block, we select the centroid pixel with coordinates z_i as the reference pixel, and it is transformed to the pixel $\hat{z}_i$. The pixel $\hat{z}_i$ is interpolated by using the coefficients in P_θ, and then

$$\hat{z}_i = \sum_{j=1}^{k_s^2} P_\theta(i,j)\ z_{i_j}, \tag{12}$$

where z_{i_j} is the coordinate vector of the jth neighboring pixel of the pixel z_i.

Then, based on these pixel pairs, we can have

$$\begin{pmatrix} a & b \\ -b & a \end{pmatrix} [z_1, \cdots, z_n] + \begin{pmatrix} T_x \\ T_y \end{pmatrix} \otimes e_n^T = [\hat{z}_1, \cdots, \hat{z}_n],$$

where $a = r\cos(\alpha), b = r\sin(\alpha)$. The parameters can be directly solved by using the Least Squares Error (LSE) approach, and the parameters r and θ can be then deduced from a and b. In this work, as the datum itself may have spatial misalignments even if it is manually cropped, we cannot evaluate the exact accuracy of the parameter estimation without ground truth, and hence we instead show the misalignment-amended images to demonstrate the accuracy of misalignment estimation in the experimental section.

Byproduct: Occlusion Detection. A byproduct of this formulation is that, when there exist image occlusions in the observed face image x, the ℓ_1 norm minimization of the error vector ε can also recover these areas as the pixels with large errors in ε. 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 $W^\sharp$.

C. Discussions

In this subsection, we discuss the relationship between our proposed general formulation for misalignment-robust face recognition and two related works [12][15] that also try to tackle this spatial misalignment problem.

1) Relationship with [12] using virtual samples: Shan *et al.* [12] introduced the concept of the *curse of misalignment*, and 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 [12] in several aspects: 1) the work [12] cannot handle image occlusion; 2) the work [12] cannot work under scenarios where the training images are already cropped; 3) the virtual samples essentially make the classification boundaries 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. In the experimental section, we compare our proposed algorithm with the work in [12] in the cases where the images are automatically cropped out and neither algorithm can estimate the spatial misalignments in a theoretically reasonable way.

Fig. 3. Demonstration of misalignment (translations) estimation and image reconstruction on the CMU PIE database. Original samples are displayed in the first row, the second row is the corresponding samples with random translations of 4 pixels, and the bottom row contains the reconstructed samples by our algorithm.

Fig. 4. Demonstration of misalignment (scales and rotations) estimation and image reconstruction on the CMU PIE database. Original samples are displayed in the first row, the second row is the corresponding samples with random scalings (1–4th columns) and rotations (5–8th columns), and the bottom row contains the reconstructed samples by our algorithm.

2) Relationship with the shift invariant PCA [15]: Tu *et al.* [15] 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, but not 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 discussed in the nest section; 3) similar to [12], it cannot handle the cases with image occlusions.

IV. Experiments

In this section, we systematically evaluate the effectiveness of our general formulation for misalignment-robust (MAR) face recognition, and we take two popular subspace learning algorithms, LDA [3] and MFA [21], as examples for the evaluation. The evaluation consists of four aspects: 1) spatial misalignment estimation and image reconstruction, 2) occlusion detection and recovery, 3) face recognition on testing data with synthesized spatial misalignments or image occlusions, and 4) face recognition under the scenario with automatic image cropping.

A. Data Sets

Four benchmark face databases, ORL, CMU PIE, YALE [1], 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, 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 this database, we randomly select half of the images of each person for model training, and the left half for testing. Since manually cropped faces are not available for FRGC database, the face images are automatically cropped and then normalized to the size of 32-by-32 pixels in the experiments.

[1]Available at http://www.face-rec.org/databases/.

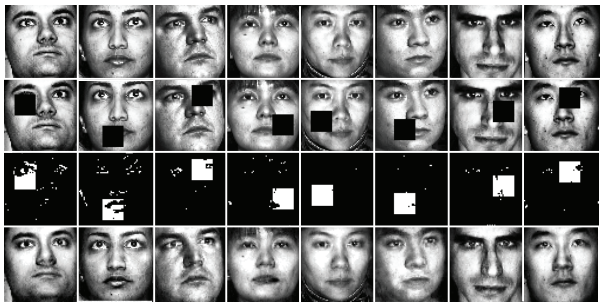

Fig. 5. Demonstration of occlusion detection on the CMU PIE database. Original samples are displayed in the first row. An 18-by-18 occlusion is randomly generated as shown in the second row. The third row shows the error maps derived from our algorithm, and the recovered images are demonstrated in the bottom row.

B. Spatial Misalignment Estimation

In this subsection, we demonstrate the spatial misalignment estimation performance of our proposed MAR formulation. The CMU PIE database is used for this evaluation. We randomly select four images per subject for model training, and from the remaining images we randomly select 8 probe images, as illustrated in the first row of Figure 3. To simulate misalignment in real cases, a random translation of $+4$ or -4 pixels in the vertical or horizontal direction is added to each probe image. Using the proposed MAR formulation, the random translations can be detected and estimated by examining the parameter vector P_θ. The images can also be reconstructed using Eqn. (7) by removing the noise term. The reconstruction results are shown in the 3rd row of Figure 3. The subspace learning algorithm used in this evaluation is LDA, and the dimension of the LDA subspace is set to 50. For better visualization, the images are normalized to the size of 64-by-64 pixels and the dimension of the principal

complementary space is set to $N - d_l$, where N is the sample number and d_l $(= 50)$ is the dimension of the LDA subspace. Also, the reconstruction performance for image scaling and rotation is demonstrated in Figure 4. For scaling, the original image is randomly scaled by a factor $r \in [0.9, 1.1]$, and for image rotation, a random rotation $\alpha \in [-10^o, +10^o]$ is imposed on the original image.

We also conducted the experiments to quantitatively evaluate the accuracies on image misalignment estimation based on our proposed algorithm. The CMU PIE database is used for the experiments and six images each person without image misalignments are used for model training. The image size and the detailed subspace learning algorithm are set the same as in the experiments for qualitative evaluation. In the evaluation, each type of misalignment is evaluated independently and all the test images are used for experiments. The image translation is set as integer within $[-6,\ 6]$ pixels for both horizontal and vertical directions, the rotation is set randomly within $(-12^o, +12^o)$, and the scaling is set randomly within $(0.85,\ 1.15]$. The detailed results on the average estimation errors within different misalignment ranges are summarized in Table I, from which we can observe: 1) the image translation estimation in both horizontal and vertical directions shows to be very satisfying; 2) for image rotation estimation, the accuracy within the range of $(-6^o, 6^o)$ is acceptable; 3) for image scaling estimation, the performance for scaling down the size is generally better than that from scaling up the size; and 4) when the misalignment range is further enlarged, the estimation error shall further increase but our proposed algorithm still shows effective within the ranges we evaluated.

C. Occlusion Detection

For facial images with occlusions, the occluded parts can be revealed by detecting the elements of ε 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 values of ε as the occluded pixels. For real images with occlusions, the occlusion area can be selected by setting an empirical threshold for the ε value to determine whether a pixel is occluded. Eight images are randomly selected from the probe set and the occlusion detection results are shown in Figure 5, 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. (7), which is demonstrated in Figure 5. The configuration for the subspace learning algorithm is the same as that for the spatial misalignment estimation.

D. 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 '$NxTy$' 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. To simulate the spatial misalignments, translation randomly generated within the interval $[-1,\ +1]$ pixels, random rotation within the interval $[-10^o,\ +10^o]$, and random scaling within $[0.9, 1.1]$ are added to the probe images. We also use the mixed spatial misalignments to simulate the misalignments brought by the automatic face alignment process. In the mixed spatial misalignment configuration, a rotation $\alpha \in [-5^o, +5^o]$, a scaling $r \in [0.95, 1.05]$, a horizontal shift $T_x \in [-1, +1]$, and a vertical shift $T_y \in [-1, +1]$ are randomly added to the original image. The detailed results with random translations for the testing data are listed in Table II, the recognition results with random scalings are shown in Table III, the recognition results with random rotations are shown in Table IV, and the performance with mixed spatial misalignments is demonstrated in Table V compared with the performance on manually cropped images. From these tables, 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 spatial misalignments are more serious in the YALE database even though they are manually cropped.

E. Face Recognition with Occlusions

In this subsection, we evaluate the face recognition performance of MAR formulation under scenarios with image

TABLE I

QUANTITATIVE IMAGE MISALIGNMENT ESTIMATION ERRORS FOR FOUR DIFFERENT TYPES OF MISALIGNMENTS EVALUATED ON THE CMU PIE DATASET.

Type	Translation T_x (pixel)					Translation T_y (pixel)				
Ground Truth	-6:-4	-3:-2	-1:1	2:3	4:6	-6:-4	-3:-2	-1:1	2:3	4:6
Average Estimation Error	0.81	0.56	0.57	0.57	0.63	0.54	0.34	0.45	0.45	0.79

Type	Rotation α (^{o})					Scaling r			
Ground Truth	(-12, -6]	(-6, -2]	(-2, 2]	(2, 6]	(6, 12)	(0.85, 0.95]	[0.95 1]	(1, 1.05]	(1.05, 1.15)
Average Estimation Error	3.49	1.50	0.486	1.11	3.02	0.0090	0.0076	0.0197	0.0236

TABLE II

RECOGNITION ACCURACY RATES (%) OF DIFFERENT ALGORITHMS ON THE THREE DATABASES WITH RANDOM IMAGE TRANSLATIONS. NOTE THAT THE BOLD NUMBERS ARE THE BEST ACCURACIES FOR EACH CONFIGURATION.

Configuration	Baselines		LDA Related Algorithms		MFA Related Algorithms	
YALE	w/o DR	PCA	Ori-LDA	MAR-LDA	Ori-MFA	MAR-MFA
N6T5	68.0	69.3	77.3	**80.0**	78.7	**80.0**
N5T6	62.2	64.4	66.7	**74.4**	70.0	73.3
N4T7	63.8	67.6	67.6	**77.1**	65.7	71.4
ORL	w/o DR	PCA	Ori-LDA	MAR-LDA	Ori-MFA	MAR-MFA
N4T6	66.3	68.8	68.3	77.1	71.3	**77.5**
N3T7	61.4	64.3	66.8	72.1	67.9	**72.5**
N2T8	56.9	55.3	54.7	62.5	54.7	**63.1**
PIE	w/o DR	PCA	Ori-LDA	MAR-LDA	Ori-MFA	MAR-MFA
N4T6	69.1	73.6	74.9	82.0	75.7	**87.8**
N3T7	66.0	70.1	67.4	82.1	68.9	**83.2**
N2T8	60.5	64.4	60.1	76.4	64.3	**76.8**

TABLE III

RECOGNITION ACCURACY RATES (%) OF DIFFERENT ALGORITHMS ON THE THREE DATABASES WITH RANDOM IMAGE SCALING.

Configure	Baselines		LDA Related Algorithms		MFA Related Algorithms	
YALE	w/o DR	PCA	Ori-LDA	MAR-LDA	Ori-MFA	MAR-MFA
N6T5	69.3	66.7	76.0	81.3	77.3	**84.0**
N5T6	65.6	65.6	74.4	**78.9**	74.4	76.7
N4T7	66.7	64.8	71.4	**77.1**	67.6	**77.1**
ORL	w/o DR	PCA	Ori-LDA	MAR-LDA	Ori-MFA	MAR-MFA
N4T6	70.4	70.0	65.0	**74.6**	65.8	73.8
N3T7	67.1	67.9	64.6	**77.1**	65.4	76.1
N2T8	56.9	57.2	54.7	60.6	55.9	**61.6**
PIE	w/o DR	PCA	Ori-LDA	MAR-LDA	Ori-MFA	MAR-MFA
N4T6	81.2	83.3	83.9	**91.8**	84.7	91.3
N3T7	74.4	76.4	81.2	87.3	81.4	**88.9**
N2T8	70.6	69.6	74.2	84.9	74.0	**85.7**

TABLE IV

RECOGNITION ACCURACY RATES (%) OF DIFFERENT ALGORITHMS ON THE THREE DATABASES WITH RANDOM IMAGE ROTATIONS.

Configuration	Baselines		LDA Related Algorithms		MFA Related Algorithms	
YALE	w/o DR	PCA	Ori-LDA	MAR-LDA	Ori-MFA	MAR-MFA
N6T5	70.7	68.0	76.0	**80.0**	76.0	78.7
N5T6	64.4	63.3	71.1	74.4	70.0	**75.6**
N4T7	67.6	66.7	70.5	73.3	70.5	**74.3**
ORL	w/o DR	PCA	Ori-LDA	MAR-LDA	Ori-MFA	MAR-MFA
N4T6	78.3	79.2	74.6	**80.0**	75.8	77.5
N3T7	73.2	73.9	71.4	78.2	73.2	**79.3**
N2T8	60.3	61.3	64.1	64.7	65.3	**65.6**
PIE	w/o DR	PCA	Ori-LDA	MAR-LDA	Ori-MFA	MAR-MFA
N4T6	81.8	84.7	77.3	86.2	79.1	**87.8**
N3T7	76.4	78.9	70.1	80.7	72.6	**83.0**
N2T8	71.2	70.6	65.9	80.4	66.3	**81.8**

TABLE V

RECOGNITION ACCURACY RATES (%) OF DIFFERENT ALGORITHMS ON THE THREE DATABASES: MANUALLY ALIGNED IMAGES VS. IMAGES WITH MIXED MISALIGNMENTS.

Configuration	Baselines		LDA Related Algorithms		MFA Related Algorithms	
YALE	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	**94.6**/78.7	90.7/68.0	93.3/**81.3**
N5T6	72.2/52.2	72.2/53.3	82.2/63.3	90.0/**73.3**	82.2/62.2	**92.2**/72.2
N4T7	72.4/54.3	72.4/53.3	82.9/61.0	**88.6/75.2**	83.8/61.9	**88.6**/73.3
ORL	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	**90.0/69.6**
N3T7	81.4/52.9	81.8/53.9	84.3/50.4	85.0/**65.7**	83.6/48.9	**86.1**/64.6
N2T8	71.6/46.9	68.8/49.1	71.3/45.3	75.6/54.4	72.2/45.9	**76.3/55.0**
PIE	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	**94.4/79.6**
N3T7	80.7/54.2	83.5/55.9	94.1/50.3	94.1/72.6	**95.0**/51.5	**95.0/73.7**
N2T8	78.8/51.8	81.8/51.6	84.7/46.2	89.1/69.6	86.7/46.4	**89.9/70.8**

TABLE VI

RECOGNITION ACCURACY RATES (%) OF DIFFERENT ALGORITHMS ON THE THREE DATABASES WITH RANDOM IMAGE OCCLUSIONS.

Configuration	Baselines		LDA Related Algorithms		MFA Related Algorithms	
YALE	w/o DR	PCA	Ori-LDA	MAR-LDA	Ori-MFA	MAR-MFA
N6T5	78.7	82.7	87.8	90.2	90.5	**92.3**
N5T6	74.4	71.1	88.7	91.6	89.6	**92.1**
N4T7	74.3	73.3	77.8	86.1	79.4	**87.1**
ORL	w/o DR	PCA	Ori-LDA	MAR-LDA	Ori-MFA	MAR-MFA
N4T6	87.1	86.3	84.6	**92.1**	86.8	91.3
N3T7	82.5	82.1	81.4	85.4	83.2	**86.1**
N2T8	71.6	69.4	69.7	71.6	70.0	**71.9**
PIE	w/o DR	PCA	Ori-LDA	MAR-LDA	Ori-MFA	MAR-MFA
N4T6	84.1	87.3	87.8	90.2	90.5	**92.3**
N3T7	80.3	82.1	88.7	91.6	89.6	**92.1**
N2T8	78.0	78.6	77.8	86.1	79.4	**87.1**

TABLE VII

RECOGNITION ACCURACY RATES (%) OF DIFFERENT ALGORITHMS ON AUTOMATICALLY CROPPED IMAGES (FOR BOTH TRAINING AND TESTING DATA).

Configuration	Baselines		LDA Related Algorithms			MFA Related Algorithms	
FRGC	w/o DR	PCA	Work [12]	Ori-LDA	MAR-LDA	Ori-MFA	MAR-MFA
50%:50%	57.4	57.5	87.2	86.0	**90.0**	86.0	89.5
YALE	w/o DR	PCA	Work [12]	Ori-LDA	MAR-LDA	Ori-MFA	MAR-MFA
N6T5	80.0	78.7	84.0	**89.3**	**89.3**	**85.3**	**85.3**
N5T6	72.2	68.9	80.0	78.9	**82.2**	78.9	**82.2**
N4T7	74.3	70.5	81.9	79.1	**84.8**	80.0	83.8

occlusions. The configuration is similar to the case with spatial misalignments, except that a 6-by-6 occlusion patch is generated at a random position for each probe image. The face recognition results are listed in Table VI, from which we can see that with the aid of occlusion detection and reconstruction, the recognition rates are generally boosted by 3–9 points from our MAR formulation, and the improvement is more dramatic when the training sample number is small.

F. Scenario with Automatic Cropping

Finally, we examine the performance of the MAR framework under the scenario with automatic cropping. We utilize the Active Shape Model [6] 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 is 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.

The algorithm introduced in [12] is also evaluated under this scenario, and 25 virtual samples are synthesized for each training sample by translating the positions of two eyes with ± 1 pixels in both directions. The detailed results are listed in Table VII, 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, compared with the scenario with manual cropping; 2) the MAR framework can greatly compensate

the effect of spatial misalignments caused by the automatic cropping process; and 3) the work in [12] can generally enhance the algorithmic robustness to image misalignments, but its improvement is not as significant as that from the MAR formulation proposed in this paper.

V. Conclusions and Future Work

In this paper, a general ℓ_1 norm minimization formulation has been proposed to provide misalignment-robust face recognition 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 can also be detected based on the ℓ_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. To the best of our knowledge, this is the first work to study the problem of general subspace-based face recognition with the consideration of both spatial misalignments and image occlusions. In the future, we plan to take pose variation as a type of specific spatial misalignment, and investigate a more general formulation to handle face recognition with pose variations. Also the extension of this work to tensor data [24] is another interesting direction for study.

VI. Acknowledgement

This work is supported by NRF/IDM Program, under research Grant NRF2008IDM-IDM004-029.

References

[1] http://www.acm.caltech.edu/l1magic/
[2] P. Phillips, P. Flynn, T. Scruggs, K. Bowyer, J. Chang, K. Hoffman, J. Marques, J. Min, and W.Worek. Overview of the Face Recognition Grand Challenge. *CVPR*, 2005.
[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.
[4] D. Cai, X. He, and J. Han. Spectral Regression: A Unified Approach for Sparse Subspace Learning. *ICDM*, pp. 73-82, 2007.
[5] H. Chen, H. Chang, and T. Liu, Local Discriminant Embedding and Its Variants. *CVPR*, vol. 2, pp. 846–852, 2005.
[6] T. Cootes, G. Edwards, and C. Taylor. Comparing Active Shape Models with Active Appearance Models. *Proceedings of the British Machine Vision Conference*, vol. 1, pp. 173-182, 1999.
[7] K. Fukunaga. *Introduction to Statistical Pattern Recognition*. Academic Press, second edition, 1991.
[8] X. He, S. Yan, Y. Hu, P. Niyogi, and H. Zhang. Face Recognition Using Laplacianfaces. *PAMI*, vol. 27, no. 3, 2005.
[9] I. Joliffe. Principal Component Analysis. *Springer-Verlag*, 1986.
[10] A. Martinez. Recognizing Imprecisely Localized, Partially Occluded and Expression Variant Faces from a Single Sample per Class. *TPAM*, vol. 24, no. 6, pp. 748-763, 2002.
[11] S. Prince and J. Elder. Probabilistic Linear Discriminant Analysis for Inferences About Identity. *CVPR*, 2007.
[12] 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.
[13] S. Shan, W. Gao, Y. Chang, B. Cao, and P. Yang. Review the strength of Gabor features for face recognition from the angle of its robustness to mis-alignment. *ICPR*, pp. 338-341, 2004.
[14] D. Swets and J. Weng. Using Discriminant Eigenfeatures for Image Retrieval. *TPAMI*, vol. 18, no. 8, pp. 891-896, 1996.
[15] 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.
[16] M. Turk and A. Pentland. Face Recognition Using Eigenfaces. *CVPR*, pp. 586–591, 1991.
[17] P. Wang, M. Green, Q. Ji, and J. Wayman. Automatic Eye Detection and Its Validation. *CVPR*, vol. 3, pp. 164-171, 2005.
[18] X. Wang and X. Tang. Random Sampling for Subspace Face Recognition. *IJCV*, vol. 79, no. 1, pp. 91-104, 2006.
[19] X. Wang and X. Tang. A Unified Framework for Subspace Face Recognition. *PAMI*, vol. 26, no. 9, pp. 1222-1228, 2004.
[20] J. Wright, A. Ganesh, A. Yang, and Y. Ma. *Robust face recognition via Sparse Representation*. TPAMI, vol. 31, no. 2, pp. 210-227, 2009.
[21] S. Yan, D. Xu, B. Zhang, H. Zhang, Q. Yang and S. Lin. Graph Embedding and Extensions: A General Framework for Dimensionality Reduction. *TPAMI*, vol. 29, no. 1, pp. 40-51, 2007.
[22] M. Yang. Kernel Eigenfaces vs. Kernel Fisherfaces: Face Recognition Using Kernel Methods. *FGR*, pp. 215-220, 2002.
[23] 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.
[24] J. Ye, R. Janardan, and Q. Li. Two-Dimensional Linear Discriminant Analysis. *NIPS*, 2004.
[25] J. Ye and T. Xiong. Null Space Versus Orthogonal Linear Discriminant Analysis. *ICML*, pp.1073–1080, 2006.
[26] W. Zhao, R. Chellappa, and A. Krishnaswamy. Discriminant Analysis of Principal Components for Face Recognition. *FGR*, pp. 336-341, 1998.