Computer Vision and Image Understanding 116 (2012) 1095–1110

Contents lists available at SciVerse ScienceDirect

Computer Vision and Image Understanding

journal homepage: www.elsevier.com/locate/cviu

Robust pose invariant face recognition using coupled latent space discriminant analysis ☆

Abhishek Sharma *, Murad Al Haj, Jonghyun Choi, Larry S. Davis, David W. Jacobs

Institute for Advanced Computer Studies, University of Maryland, College Park, MD 20742, United States

ARTICLE INFO

Article history:
Received 25 January 2012
Accepted 4 August 2012
Available online 14 August 2012

Keywords:
Pose-invariant-face recognition
Coupled latent space
PLS
CCA
Discriminant coupled subspaces

ABSTRACT

We propose a novel pose-invariant face recognition approach which we call Discriminant Multiple Coupled Latent Subspace framework. It finds the sets of projection directions for different poses such that the projected images of the same subject in different poses are maximally correlated in the latent space. Discriminant analysis with artificially simulated pose errors in the latent space makes it robust to small pose errors caused due to a subject's incorrect pose estimation. We do a comparative analysis of three popular latent space learning approaches: Partial Least Squares (PLSs), Bilinear Model (BLM) and Canonical Correlational Analysis (CCA) in the proposed coupled latent subspace framework. We experimentally demonstrate that using more than two poses simultaneously with CCA results in better performance. We report state-of-the-art results for pose-invariant face recognition on CMU PIE and FERET and comparable results on MultiPIE when using only four fiducial points for alignment and intensity features.

1. Introduction

Face recognition is a very challenging problem due to variations in pose, illumination and expression. Research in this area spans a wide range of statistical and geometric pattern recognition algorithms for tackling the aforementioned difficulties. Most successful face recognition approaches require accurate alignment and feature correspondence between the face images to be compared. However, in many real-life scenarios, face images appear in different poses causing correspondence problem. There has been a large body of work dealing with pose variation, but still fast and accurate recognition is a challenge. For a comprehensive and recent survey of pose invariant face-recognition please refer to [2,1].

We can regard a face image as a vector in $\Re^D$. The coordinate axes defined for each pixel will constitute a *representation scheme* ($\mathcal{S}$) for the face which is basically the set of column vectors of an identity matrix in $\Re^D$ space. Corresponding pixels across different subjects' faces roughly correspond to the same facial region in the absence of pose difference. This *feature correspondence* facilitates comparison. In fact, feature correspondence is essential for comparison based on a learned model. For faces especially, it has been shown to be crucial [3]. Unfortunately, face images under different poses lose the feature correspondences because of missing facial regions, unequal dimensions and/or *region displacements*. Region displacement refers to the same facial region at different indices in feature vectors (see Fig. 1).

We propose to obtain pose-specific representation schemes $\mathcal{S}_i$'s so that the projection of face vectors onto the appropriate representation scheme will lead to correspondence in the common projected space, which facilitates direct comparison. A representation scheme can also be regarded as a collection of projection directions, which we refer to as a *projector*. Intuitively, projectors are feature extractors through which the common information from multiple poses is collected and transferred to a common representation scheme which we term as *latent space*. Given a set of projectors $\mathcal{S}_p$ and $\mathcal{S}_q$ for gallery pose p and probe pose q. $\mathcal{S}_p$ and $\mathcal{S}_q$ can be used to project the gallery and probe images to the latent space where direct comparison can be done due to feature correspondence. The pose-specific projectors and associated latent space taken together are termed as *Correspondence Latent Subspace* or CLS because projection into the latent space provides correspondence.

In a preliminary version of the paper [11], we showed the conditions under which such latent spaces exist and used Partial Least Square (PLS) [20,21,23,22] to obtain them. PLS has been used before for face recognition, but it was used either as a feature extraction tool [27–30] or a classifier [31]. In contrast to the previous approaches, we used PLS to learn sets of CLS for different pose-pairs to facilitate pose-invariant face recognition. Our work shows that linear projection to latent space is an effective solution for pose-invariant face recognition, which is considered to be a highly non-linear problem [12,42,25,44]. Working independently, authors in [26] have also used PLS for learning sets of CLS for different pose-pairs. However, they have used Gabor features and probabilistic

☆ This paper has been recommended for acceptance by K.W. Bowyer.

* Corresponding author.
E-mail address: abhisharayiya@gmail.com (A. Sharma).

1077-3142/$ - see front matter © 2012 Elsevier Inc. All rights reserved.
http://dx.doi.org/10.1016/j.cviu.2012.08.001

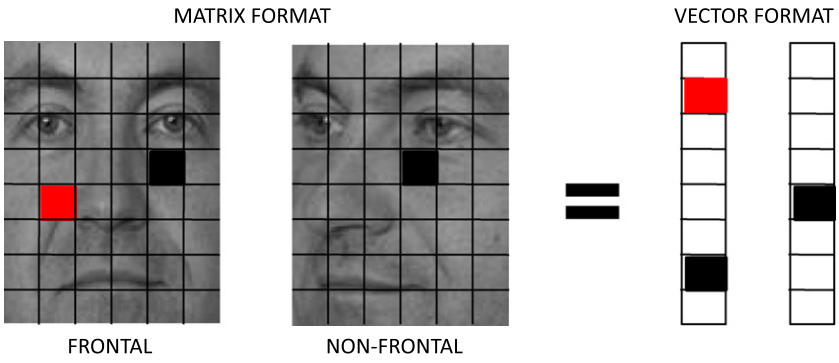


Fig. 1. An example showing lack of correspondence due to missing regions and region displacement for pose variation. Black and red blocks indicate region displacement and missing region, respectively. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

fusion of local scores for the final decision. Unlike our approach, they have not considered learning a common CLS for multiple poses. Surprisingly, our approach with simple intensity features outperforms previous work and gives state-of-the-art results on CMU PIE data for pose invariant face recognition and comparable results on FERET and MultiPIE.

Based on the general correspondence latent space model, the pose-invariant face recognition problem reduces to estimation of the gallery and probe pose and obtaining an appropriate CLS for recognition. We require a training set of face images in gallery and probe poses to learn CLS for these poses. In this work, we assume prior knowledge of *ground-truth* gallery and probe poses. Ground-truth pose refers to the pose which is reported in the dataset. We also require training data with face images roughly in gallery and probe poses. The subject identities in the training and testing data are different and mutually exclusive. These assumptions are quite standard for learning based methods and have been used by many researchers in the past [15,16,42,25,37,43,14,13,9,26].

Our previous, simple PLS based framework [11] worked well for CMU PIE dataset, which contains face images in a tightly controlled acquisition scenario that ensures that the ground-truth poses are very close to the actual poses. But it did not perform as expected on less controlled and larger datasets, e.g. FERET [17] and MultiPIE [33]. On one hand, larger gallery size requires more discriminative features for classification but our previous approach is generative and does not use label information to learn projectors that are discriminative. On the other hand, a less controlled acquisition scenario gives rise to *pose errors*, which refers to the situation where the actual pose of the face image differs from the projector learned for that pose. Even if the difference is small (generally around ±10°), it can cause loss of correspondence which degrades the performance. Pose errors can be caused due to wrong pose estimation or head movement at the time of acquisition. The presence of pose errors is supported from significant differences between the estimated poses [4] and the ground-truth poses [6] for FERET dataset and our own experiments to estimate pose (Section 4.5).

In order to make our framework practically applicable we need to account for large gallery sizes and pose errors. Therefore, we extend our original PLS framework [11] to a two-stage framework for pose invariant face recognition. The first stage learns pose-specific representation schemes for gallery/probe pose pairs (which we assume to be known beforehand) using a training set that has face images in roughly the same poses. The second stage learns discriminative directions in the Correspondence Latent Subspace (CLS) that has three added advantages:

- Providing an identity based discriminative representation which is known to outperform generative representation [10].
- Achieving insensitivity to pose errors that are present in real-life as well as controlled scenarios.
- Exploiting multiple face samples per person in different poses for supervised learning, which was otherwise not possible due to modality difference.

We empirically noticed the improvement in recognition accuracy due to all these factors in the overall performance and report state-of-the-art pose recognition results for 2D based methods on CMU PIE and FERET and comparable to best published results on MultiPIE. A theoretical and empirical comparison between three popular methods CCA, PLS and BLM for learning the CLS is done under different scenarios. We also provide our hand-annotated fiducial points for FERET and MultiPIE publicly available on our website (http://www.umiacs.umd.edu/bhokaal/data/FERET_MultiPIE_fiducials.tar) to promote research with these datasets.

This is an extended version of our conference paper [11]. The original conference version does not include the second stage discriminative learning and the results on FERET and MultiPIE. However, the conference version had a more detailed explanation of PLS which we omit here due to space constraints.

The rest of the paper is organized as follows: Section 2 gives a brief review of related approaches for pose invariant face recognition, Section 3 discusses some background. Section 4 describes the proposed approach with PLS and effect of pose errors. Section 5 discusses the two-stage discriminative framework followed by experimental analysis in Section 6. Finally, we conclude and discuss salient points of the approach in Section 7.

2. Previous work on pose-invariant face recognition

In [4], the authors proposed a 3D Morphable Model (3DMM) for faces and used the fact that 3D face information extracted as *shape and texture* features remains the same across all poses. Hence, given a 2D image they estimated the corresponding 3D model and matched in the 3D shape and texture space. This method is among the best performing algorithms for pose invariant face recognition but it heavily depends on the accurate extraction of 3D information from the 2D image which itself is a difficult problem and computationally intensive, making it too slow for real-time application. It also requires 6–8 fiducial points and 3D face models during training to learn the 3D shape and texture space. Recently, Generic Elastic Models (GEMs) [38] showed that 3D depth information is not discriminative for pose invariant face recognition. Thus, a generic face depth map can be elastically deformed for a given 2D face to generate the corresponding 3D model leading to a fast version of 3DMM (2–3 s per image). They also extracted all the required 79 fiducial landmarks automatically. The 3D pose normalization

approach presented in [5] synthesizes a virtual frontal view and then extracts Local Gabor Binary Patterns (LGBP) [52] to find the closest match in the gallery. It can handle continuous variation in pose and has impressive performance on different datasets for ±45° in pitch and ±30° yaw variation. A different 3D geometric approach is based on stereo matching [35,36] which uses four fiducial points to obtain the epipolar geometry and dynamic programming to match corresponding pixels. This approach has shown impressive performance on CMU PIE data set.

Locally Linear Regression or LLR [12] uses face images/patches to be the bases of the representation scheme, assuming that a face image/patch in pose p can be faithfully represented as a linear combination of a set of face images/patches and that the coefficients of linear combinations remain roughly constant across different poses. The coefficients of combination were learned using linear regression. Recently, [42] has reported significantly improved performance by using Ridge regression to estimate coefficients of a linear combination of a subject's face image in terms of training subject's images in the same pose and comparing the coefficients using normalized correlation. They have used Gabor features [41] at five hand-clicked facial points rather than simple pixel intensity to further enhance the performance. Similarly, the associate-predict model [44] divides face images into patches and extracts LBP [39], SIFT [40], Gabor [41] and Learning based descriptors (LE) [49] as features. Then each patch is associated with a similar patch from a set of generic face images under approximately the same pose (*associate* step). In the prediction step, the associated patch's corresponding patch in the gallery pose is used as a proxy for the original patch for matching purposes. All the above-mentioned approaches are essentially 2D approximations of the 3DMM theory which is not always correct. The strength of the approximation relies heavily on the validity of the key assumption that the coefficients across pose remain almost the same. We argue that it may not hold for 2D face images unless it is forced explicitly [11,9]. In [13], the authors realized this shortcoming and used Canonical Correlational Analysis (CCA) [19] to learn a pair of subspaces which make the projected images in the latent space maximally correlated. They also used a region based discriminative power map for face pixels modeled as a probability distribution [25]. We also use CCA to learn CLS but we use more than two poses simultaneously and pool information from multiple poses using latent space discriminant analysis. In [43], an attempt was made to learn the patch correspondence between frontal and non-frontal poses by using a batch version of Lucas–Kanade optical flow algorithm [45]. However, they use only two poses at a time and the discrimination is not based on label information.

TFA [15] and PLDA [16] use generative models to synthesize face images of a person across different poses from a common latent variable which they call Latent Identity Variable or LIV. At the time of recognition, the images are transformed to the LIV space using a pose-specific linear transformation and recognition is carried out in that space. The accuracy of the approach depends on the validity of the factor model in terms of modeling the problem and the quality of the learned model parameters. They use the EM algorithm [46] to learn the model parameters which is prone to local minima and computationally expensive. Moreover, the assumption that a single LIV can be used to faithfully generate all the different poses of a person seems to be over simplified and may not be true. It becomes evident from poor performance even for small poses angles with simple intensity features. To improve the performance, they used 14 hand clicked points on face images to extract Gabor filter response which are more discriminative than raw pixels. But accurate location of fiducial-points in non-frontal images is still an open problem. A related patch-whole approach was proposed in [14] which tries to model the differential distribution of a gallery image patch and the whole probe face. The advantage of this approach lies in the fact that due to a patch-whole matching scheme it is comparatively robust to small pose-estimation errors. In the next section we discuss some relevant literature for learning CLS.

3. Background

In this section we discuss the details of Bilinear Model (BLM), Canonical Correlational Analysis (CCA) and Partial Least Square (PLS) because we need them later on. All of these methods find a set of representation schemes which make the projected images of the same person *similar* to each other in the latent space. The definition of *similar* varies with the method; for instance, CCA makes them maximally correlated while PLS maximizes the covariance between them. We also draw a theoretical comparison between these approaches.

Notation: Throughout the paper, superscripts denote indexing across identity, subscript denotes modality/pose, vectors are denoted as straight bold small alphabets ($\mathbf{x}$), variable/constants as small italic alphabets (a) and matrices as capital italic letters (A). Hence, the face image of i^{th} person in pose p is denoted as $\mathbf{x}_p^i$ and a matrix of face samples in pose p as X_p.

3.1. Bilinear model

Tannenbaum and Freeman [18] proposed a bilinear model for separating *style* and *content*. In pose invariant face recognition, style corresponds to pose and content corresponds to subject identity. They suggest methods for learning BLMs and using them in a variety of tasks, such as identifying the style of a new image with unfamiliar content, or generating novel images based on separate examples of the style and content. However, their approach also suggests that their content-style models can be used to obtain a style invariant content representation that can be used for classification of a sample in a different style. Following their asymmetric model, they concatenate the ith subject's images under M different modalities/poses ($\mathbf{y}_m^i : m = 1, 2, \ldots M$) to make a long vector $\mathbf{y}^i$ and construct matrix Y having columns as $\mathbf{y}^i$ with $i = \{1, 2, \ldots, N = \#\ subjects\}$ such that:

$$Y = \begin{pmatrix} \mathbf{y}_1^1 & \mathbf{y}_1^2 & \cdots & \mathbf{y}_1^N \\ \mathbf{y}_2^1 & \mathbf{y}_2^2 & \cdots & \mathbf{y}_2^N \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{y}_M^1 & \mathbf{y}_M^2 & \cdots & \mathbf{y}_M^N \end{pmatrix} = \begin{pmatrix} \mathbf{y}^1 & \mathbf{y}^2 & \ldots & \mathbf{y}^N \end{pmatrix} \tag{1}$$

Modality matrices A_m which can be thought of as different representation schemes for a CLS model can be obtained by decomposing the matrix Y using SVD as

$$Y = USV^T = (US)V^T = (A)B \tag{2}$$

A can be partitioned $A^T = \begin{pmatrix} A_1^T & A_2^T & \ldots & A_M^T \end{pmatrix}$ to give different CLS representation schemes A_m's where m represents different poses.

3.2. CCA

CCA is a technique that learns a set of M different projectors from a set of observed *content* under M different *styles*. The projections of different *styles* of a particular *content* are maximally correlated in the projected space. Hence, CCA can be used to learn a common intermediate subspace in which projections of different pose images of the same subject will be highly correlated and recognition can be done on the basis of the correlation score. Given a set of face images of N different subjects under M different poses, CCA learns a set of K dimensional subspaces $W_m = \{\mathbf{w}_m^k : \mathbf{w}_m^k \in \Re^{Dm}; k = 1, 2, \ldots K\}$ for $m = 1, 2, \ldots, M$ such that [19]:

$$\begin{pmatrix} C_{11} & C_{12} & \dots & C_{1M} \\ C_{21} & C_{22} & \dots & C_{2M} \\ \vdots & \vdots & \ddots & \vdots \\ C_{M1} & C_{M2} & \dots & C_{MM} \end{pmatrix} \begin{pmatrix} \mathbf{w}_1^k \\ \mathbf{w}_2^k \\ \vdots \\ \mathbf{w}_M^k \end{pmatrix} = (1+\lambda^k) \begin{pmatrix} C_{11} & 0 & \dots & 0 \\ 0 & C_{22} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & C_{MM} \end{pmatrix} \begin{pmatrix} \mathbf{w}_1^k \\ \mathbf{w}_2^k \\ \vdots \\ \mathbf{w}_M^k \end{pmatrix} \Rightarrow CW = W(I+\Lambda) \tag{3}$$

where Dm is the feature dimension of the mth style, $C_{ij} = \frac{1}{N} Y_i (Y_j)^T$ and Λ is a diagonal matrix of eigen-values λ^k, N is the number of training subjects and Y_i is defined in the previous sub-section. Eq. (3) is a generalized eigenvalue problem which can be solved using any standard eigensolver. The columns of the projector matrices W_m will span a linear subspace in modality m. So, when modalities are different poses we get a set of vectors spanning a linear subspace in each pose.

3.3. Partial least squares

Partial Least Square analysis [23,20–22] is a regression model that differs from Ordinary Least Square regression by first projecting the regressors (input) and responses (output) onto a low dimensional latent linear subspace. The PLS projectors try to maximize the covariance between latent scores of regressors and responses. Hence, we can use PLS to obtain CLS for two different poses in the same way as BLM and CCA.

There are several variants of PLS analysis based on the objective function and related constraints to learn the latent space, see [22] for details on different PLS algorithms. In this paper, we have used the factor model assumption given in [22,20] to develop intuitions and a variant of NIPALS given in [21] to learn the projectors.

Following the same conventions as for BLM and CCA, Y_p represents a matrix containing face images in pose p as its columns. PLS greedily finds vectors $\mathbf{w}_p$ and $\mathbf{w}_q$ such that

$$\max_{\mathbf{w}_p, \mathbf{w}_q} \left(cov\left[Y_p^T \mathbf{w}_p, Y_q^T \mathbf{w}_q \right]^2 \right) \quad s.t. \quad \|\mathbf{w}_p\| = \|\mathbf{w}_q\| = 1 \tag{4}$$

3.4. Difference between BLM, PLS and CCA

Although BLM, CCA and PLS try to achieve the same goal but the difference in their objective functions leads to different properties. BLM tries to preserve the variance present in different feature spaces and does not explicitly try to make projected samples similar. It is interesting to compare the objective function of PLS with that of CCA to emphasize the difference between the two. CCA tries to maximize the correlation between the latent scores

$$\max_{\mathbf{w}_p, \mathbf{w}_q} \left(corr\left[Y_p^T \mathbf{w}_p, Y_q^T \mathbf{w}_q \right]^2 \right) \quad s.t. \quad \|\mathbf{w}_p\| = \|\mathbf{w}_q\| = 1 \tag{5}$$

where

$$corr(\mathbf{a}, \mathbf{b}) = \frac{cov(\mathbf{a}, \mathbf{b})}{var(\mathbf{a})\, var(\mathbf{b})} \tag{6}$$

putting the expression from (6) into (4) we get the PLS objective function as:

$$\max_{\mathbf{w}_p, \mathbf{w}_q} \left(\left[var(Y_p^T \mathbf{w}_p) \right] \left[corr\left(Y_p^T \mathbf{w}_p, Y_q^T \mathbf{w}_q \right) \right]^2 \left[var\left(Y_q^T \mathbf{w}_q \right) \right] \right) \quad s.t. \quad \|\mathbf{w}_p\| = \|\mathbf{w}_q\| = 1 \tag{7}$$

It is clear from (7) that PLS tries to correlate the latent score of regressor and response as well as captures the variations present in the regressor and response space too. CCA only tries to correlate the latent score hence CCA may fail to generalize well to unseen testing points and even fails to differentiate between training samples in the latent space under some restrictive conditions. Let us consider a simplified case where PLS will succeed and both BLM and CCA will fail to obtain meaningful directions. Suppose we have two sets of 3D points X and Y and x_i^j and y_i^j denote the jth element of the i^{th} data point in X and Y. Suppose that the first coordinates of x_i and y_i are pairwisely equal and the variance of the first coordinate is very small and insufficient for differentiating different samples. The second coordinates are correlated with a correlation-coefficient $\rho \leqslant 1$ and the variance present in the second coordinate is ψ. The third coordinate is almost uncorrelated and the variance is $\gg \psi$.

$$\begin{aligned} &\forall i, x_i^1 = y_i^1 = k \quad \Rightarrow \quad var(X^1) = var(Y^1) = \alpha \ll \psi \\ &corr(X^2, Y^2) = \rho \text{ and } var(X^2), \quad var(Y^2) \approx \psi \\ &corr(X^3, Y^3) \approx 0 \text{ and } var(X^3), \quad var(Y^3) \gg \psi \end{aligned} \tag{8}$$

Under this situation CCA will give the first coordinate as the principal direction which projects all the data points in sets X and Y to a common single point in the latent space, rendering recognition impossible. BLM will find a direction which is parallel to the third coordinate, which preserves the inter-set variance but loses all the correspondence. PLS, however, will opt for the second coordinate, which preserves variance (discrimination) as well as maintains correspondence which is crucial for our task of multi-modal recognition.

One major disadvantage of PLS as compared to CCA and BLM is that the extension of PLS to more than two modalities leads to a poor set of projectors and is computationally expensive. So PLS is not suited for our Discriminant Multiple CLS framework (discussed later) which requires coupled projectors for multiple poses. On the other hand, CCA and BLM easily extend to multiple poses following (1) and (3). However, the objective function and empirical results in [11] suggest that CCA is better than BLM for cross-modal recognition. Hence, we use CCA for the purpose of learning multiple CLS.

3.5. Linear discriminant analysis

There are two kinds of variations found in data samples: within-class and between-class variation. Within-class variation refers to variation present among the samples of the same class and between-class variation refers to the variation between the samples from different classes. Ideally, for a classification task we would like that the within-class variation is minimized and between-class variation is maximized simultaneously. The quantitative measure of within-class and between-class variation are the within-class scatter matrix S_W and between-class scatter matrix S_B

$$\begin{aligned} S_W &= \sum_{i=1}^{C} \sum_{j=1}^{N_c} \left(\mathbf{x}_i^j - \mathbf{m}_i \right) \left(\mathbf{x}_i^j - \mathbf{m}_i \right)^T \\ S_B &= \sum_{i=1}^{C} (\mathbf{m}_i - \mathbf{m})(\mathbf{m}_i - \mathbf{m})^T \end{aligned} \tag{9}$$

Linear discriminant analysis or LDA tries to find a projection matrix W that maximizes the ratio of S_B and S_W

$$W_{opt} = \text{argmax}_W \frac{|W^T S_B W|}{|W^T S_W W|} \tag{10}$$

It leads to the following generalized eigen-value problem

$$S_B \mathbf{w}_i = \lambda_i S_W \mathbf{w}_i \quad i = \{1, 2, \dots C-1\} \tag{11}$$

Here, $\mathbf{x}_i^j$ is the jth sample for the ith class, $\mathbf{m}_i$ is the ith class mean, $\mathbf{m}$ is the total mean, C is the number of classes, N_c is the number of

samples for class c, λ_i's are the generalized eigen-values and $W = [\mathbf{w}_1 \quad \mathbf{w}_2 \quad \dots \quad \mathbf{w}_{C-1}]$.

4. PLS based correspondence latent space and pose error

In this section we first discuss the conditions under which CLS can account for pose difference and explain the PLS based framework for pose invariant face recognition and compare it to previous work on the CMU PIE dataset. Then we evaluate the performance of the PLS based framework on larger and less controlled datasets, e.g. FERET and MultiPIE to show that it does not perform as expected. Then, we carry out a performance drop study to understand the reason of poor performance and based on the observations we propose a novel extension of our original framework to account for the factors causing performance drop.

4.1. When CLS can account for pose

We can use a CLS framework to find linear projections that map images taken from two poses into a common subspace. However, a CLS based framework cannot be expected to lead to effective recognition when such projections do not exist. In this section, we show some conditions in which projections of images from two poses exist in which the projected images are perfectly correlated (and in fact equal). Then we show that these conditions hold for some interesting examples of pose-invariant face recognition. However, only the existence of such projections is not sufficient to guarantee good recognition performance, we must also be able to obtain them, which could be difficult or even intractable in some cases. Therefore, we will empirically assess the actual performance of the proposed approach in Section 6. In a number of cases, images taken in two poses can be viewed as different, linear transformations of a single ideal object. Let $\mathbf{i}$ and $\mathbf{j}$ denote column vectors containing the pixels of face images of the same person in two poses. We denote by r a matrix (or column vector) that contains an idealized version of $\mathbf{i}$ and $\mathbf{j}$, such that we can write:

$$\mathbf{i} = A\mathbf{r} \text{ and } \mathbf{j} = B\mathbf{r} \tag{12}$$

for some matrices A and B. We would like to know when it will be possible to find projection directions $\mathbf{p}_1$ and $\mathbf{p}_2$ that project sets of images into a 1D space in which these images are coupled. We consider a simpler case in which the projections can be made equal, i.e. when we can find $\mathbf{p}_1$ and $\mathbf{p}_2$ such that for any $\mathbf{i}$ and $\mathbf{j}$ satisfying (12) we have:

$$\begin{aligned} &\mathbf{p}_1^T\mathbf{i} = \mathbf{p}_2^T\mathbf{j} \Rightarrow \mathbf{p}_1^T A\mathbf{r} = \mathbf{p}_2^T B\mathbf{r} \\ &\mathbf{p}_1^T A = \mathbf{p}_2^T B \end{aligned} \tag{13}$$

Eq. (13) can be satisfied if and only if the row spaces of A and B intersect, as the LHS of the (13) is a linear combination of the rows of A, while the RHS is a linear combination of the rows of B. We now consider the problem that arises when comparing two images of the same 3D scene (face) taken from different viewpoints. This raises problems of finding a correspondence between pixels in the two images, as well as accounting for occlusion. To work our way up to this problem, we first consider the case in which there exists a one-to-one correspondence between pixels in the image, with no occlusion.

Permutations: In this case, we can suppose that A is the identity matrix and B is a permutation matrix, which changes the location of pixels without altering their intensities. Thus, both of A and B are full rank, and in fact they have a common row space. So, there exist $\mathbf{p}_1$ and $\mathbf{p}_2$ that will project $\mathbf{i}$ and $\mathbf{j}$ into a space where they are equal.

Stereo: We now consider a more general problem that is commonly solved by stereo matching. Suppose we represent a 3D object with a triangular mesh. Let $\mathbf{r}$ contains the intensities on all faces of the mesh that appear in either image (We will assume that each pixel contains the intensity from a single triangle. More realistic rendering models could be handled with slightly more complicated reasoning). Then, to generate images appropriately, A and B will be matrices in which each row contains one 1 and 0 otherwise. A (or B) may contain identical rows, if the same triangle projects to multiple pixels. The rank of A will be equal to the number of triangles that create intensities in $\mathbf{i}$, and similarly for B. The number of columns in both matrices will be equal to the number of triangles that appear in either image. So their row spaces will intersect, provided that the sum of their ranks is greater than or equal to the length of $\mathbf{r}$, which occurs whenever the images contain projections of any common pixels. As a toy example, we consider a small 1D stereo pair showing a dot in front of a planar background. We might have $\mathbf{i}^T = [7825]$ and $\mathbf{j}^T = [7235]$. In this example we might have $\mathbf{r}^T = [78{,}235]$ and

$$A = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix} \quad B = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix}$$

It can be inferred from the example that row spaces of A and B intersect hence we expect the CLS framework to work.

4.2. Partial least square based CLS

A PLS based framework learns projectors for every possible gallery-probe pose-pair using a training set of subjects appearing in similar gallery-probe pose-pairs. Let us denote gallery and probe poses as g and p respectively. Let X_g $(d_g \times N)$ and X_p $(d_p \times N)$ be the data matrices with columns as mean subtracted image vectors in pose g and p respectively, where d_g and d_p are gallery and probe image dimensions and N is the number of training subjects. PLS finds projectors W_g $(d_g \times K)$ and W_p $(d_g \times K)$ with K equals the number of PLS factors for pose g and p, such that

$$\begin{aligned} X_g &= W_g T_g + R_g \\ X_p &= W_p T_p + R_p \\ T_p &= D T_g + R \end{aligned} \tag{14}$$

Here, T_g $(K \times N)$ and T_p $(K \times N)$ are the latent projections of images in the CLS, R_g $(d_g \times N)$, R_p $(d_p \times N)$ and R $(K \times N)$ are residual matrices in appropriate spaces and D is a diagonal matrix that scales the latent projections of gallery images to make it equal to the probe image's projection in the latent space. Fig. 2 depicts the PLS framework pictorially. The detailed step by step algorithm to obtain these variables is given in [21].

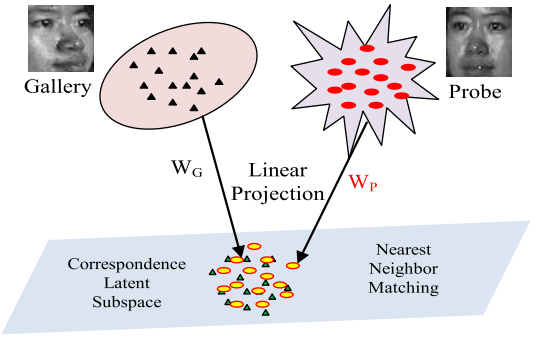


Fig. 2. PLS based framework for pose invariant face recognition, W_g and W_p are learned using PLS and training images in gallery and probe pose.

4.3. PLS on CMU PIE

The PLS based framework is used for pose invariant face recognition on CMU PIE dataset [24] which has been used by many researchers previously for evaluation. This dataset contains 68 subjects in 13 different poses and 23 different illumination conditions. We took subject IDs from 1 to 34 for training and the remaining (35–68) for testing. As we are dealing with pose variation only, we took all the images in frontal illumination which is illumination number 12. As a pre-processing step, four fiducial points (both eye's centers, nose tip and mouth) were manually annotated and an affine transformation was used to register the faces based on the fiducial points. After all the faces are aligned in corresponding poses we cropped 48×40 facial region. Images were turned into gray-scale and intensity values mapped between 0 and 1 were used as features. The number of PLS factors was set to be 30. Choosing more than 30 did not improve the performance but choosing less than 30 worsens the performance. The resulting CLS framework was termed as PLS^{30}, indicating 30 dimensional CLS obtained using PLS. The accuracy for all possible gallery-probe pairs is given in Table 1. For comparing our approach with other published works we calculated the average of all gallery-probe pairs and the resulting accuracy is listed in Table 2. Some authors have reported their results on CMU PIE data with only frontal pose as gallery and a subset of non-frontal poses as probe. For comparison we also list the gallery and probe setting in Table 2. Ridge + (Intensity/Gabor) refers to the approach of [42] with raw intensity and Gabor filter response (with probabilistic local score fusion) at fiducial locations as feature, respectively. Similarly, PLS-(Holistic/Gabors) refers to the use of PLS to learn coupled latent space with raw intensity feature from the whole face and probabilistic fusion of local scores based on Gabor filter response at fiducial locations, respectively. A simple comparison clearly reveals that PLS^{30} approach outperforms all the methods. It should be noted that the comparison with 3DMM + LGBP [5] is not fair because the results in [5] are reported on 67 subject gallery whereas, we report on 34 subject gallery. However, we still include it for the sake of completeness.

4.4. Performance drop on FERET and MultiPIE

In this subsection, we first show the results of PLS based framework on FERET and MultiPIE datasets and discuss the reason behind the poor performance. Subsequently, we propose our extended two-stage discriminative approach followed by a detailed analysis of model parameters on the overall performance.

The performance of PLS based approach on two larger and less-controlled datasets (FERET and MultiPIE) is shown in Fig. 9a and b, respectively. From the figures it is evident that performance has decreased significantly for both MultiPIE and FERET. The most obvious reason is the increased number of testing subjects (gallery); FERET and MultiPIE have almost 3 and 7 times as many testing subjects as compared to CMU PIE, respectively. As the number of testing subjects increases, we need a discriminative representation for effective classification. All three, i.e. CCA, BLM and PLS are generative in nature, hence, the decline in accuracy with increasing number of testing subject is natural. Secondly, we noticed that some of the faces in the dataset were off by a few degrees from the reported pose in the dataset. Especially for FERET, [4] has reported estimated poses which are very different from the ground-truth poses supplied with the dataset. Since projectors are learned using training images from FERET and MultiPIE, this leads to pose difference between the projectors and images. We term this phenomenon as *pose error*. It can occur because of head movement during acquisition or wrong pose estimation. Suppose, we learn two projectors for a 0°/30° gallery/probe pose pair. Let us assume that the 30° testing images are not actually 30° but $(30 \pm \theta)^\circ$ with $\theta \in [0, 15]$. For $\theta \leqslant 5$, the projectors and the testing images will have sufficient pixel correspondence. But for $\theta \geqslant 5$, we face the loss of correspondence, resulting in poor performance. Pose errors are inevitable and present in real-life as well as controlled conditions which is evident from FERET and MultiPIE. Moreover, due to different facial structures we may expect loss of correspondence for pose angles greater than 45°. For example, both the eyes of Asians are visible even at a pose angle of around 60° because of relatively flat facial structure as compared to European or Caucasian for which the second eye becomes partially or totally occluded at 60°. This leads to missing facial regions at large pose angles which creates loss of correspondence. These pose errors become more frequent and prominent with increasing pose angles.

4.5. Pose estimation

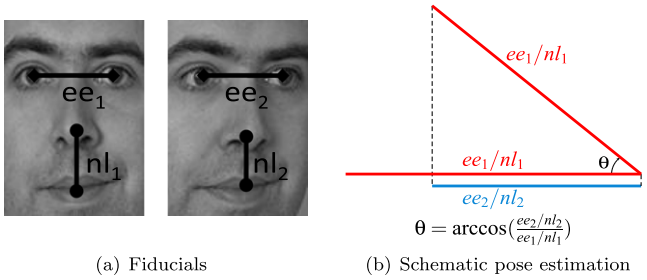
In order to show that the poses provided in the FERET and MultiPIE databases are inaccurate, we assume that for each subject the frontal pose is correct and use this information to estimate the non-frontal poses; the change in the distance between the eyes of the subject, with respect to the distance in frontal pose, is used to calculate the new pose. In general, the change in the observed eye distance can be due to two factors: change in pose and/or change in the distance between the camera and the face. For the change in the face-camera position, the distance between the nose and the lip can be used to correct this motion, if present. For the pose change, in the two datasets, there is negligible change in yaw and the Euclidean distance automatically correct for any roll change, i.e. in-plane rotation; therefore, the Euclidean eye distance once corrected by the nose-lip distance can be directly used to measure the pitch pose.

Table 1
CMU PIE accuracy using 1-NN matching and PLS with 30 dimensional CLS overall accuracy is **90.08**.

Probe→ Gallery↓	c34 −60, 0	c31 −45, 15	c14 −45, 0	c11 −30, 0	c29 −15, 0	c09 0, 15	c27 0, 0	c07 0, 0	c05 15, 0	c37 30, 0	c25 45, 0	c02 45, 15	c22 60, 0	**Avg**
c34	−/−	88.0	94.0	94.0	91.0	88.0	91.0	97.0	85.0	88.0	70.0	85.0	61.0	**86.2**
c31	85.0	−/−	100.0	100.0	100.0	88.0	85.0	91.0	85.0	88.0	76.0	85.0	76.0	**88.4**
c14	97.0	100.0	−/−	100.0	97.0	91.0	97.0	100.0	91.0	100.0	82.0	91.0	67.0	**92.8**
c11	79.0	97.0	100.0	−/−	100.0	88.0	100.0	100.0	97.0	97.0	85.0	88.0	67.0	**91.6**
c29	76.0	94.0	100.0	100.0	−/−	100.0	100.0	100.0	100.0	100.0	85.0	91.0	73.0	**93.3**
c09	76.0	88.0	91.0	94.0	94.0	−/−	97.0	94.0	91.0	88.0	82.0	79.0	70.0	**87.2**
c27	85.0	91.0	97.0	100.0	100.0	100.0	−/−	100.0	100.0	100.0	85.0	88.0	79.0	**93.9**
c07	79.0	91.0	97.0	100.0	100.0	97.0	100.0	−/−	100.0	97.0	85.0	91.0	76.0	**92.9**
c05	79.0	97.0	97.0	94.0	100.0	94.0	100.0	100.0	−/−	97.0	91.0	91.0	82.0	**93.6**
c37	79.0	94.0	100.0	94.0	94.0	88.0	94.0	94.0	97.0	−/−	100.0	100.0	94.0	**94.1**
c25	67.0	82.0	76.0	79.0	88.0	88.0	88.0	91.0	94.0	97.0	−/−	97.0	76.0	**85.5**
c02	76.0	88.0	88.0	94.0	94.0	88.0	97.0	94.0	100.0	100.0	100.0	−/−	97.0	**93.1**
c22	64.0	70.0	64.0	79.0	76.0	67.0	82.0	82.0	85.0	91.0	85.0	91.0	−/−	**78.4**

Table 2
Comparison of PLS with other published work on CMU PIE.

Method	Gallery/Probe	Accuracy	PLS^{30}
Eigenface [37]	All/all	16.6	**90.1**
ELF [37]	All/all	66.3	**90.1**
FaceIt [37]	All/all	24.3	**90.1**
4ptSMD [35]	All/all	86.8	**90.1**
SlantSMD [36]	All/all	**90.1**	**90.1**
Ridge + Intensity [42]	c27/rest all	88.24	**93.9**
PLS-Holistic [26]	c27/rest all	81.44	**93.9**
Yamada [25]	c27/rest all	85.6	**93.9**
LLR [12]	c27/c (05, 07, 09, 11, 37, 29)	94.6	**100**
PGFR [48]	c27/c (05, 37, 25, 22, 29, 11, 14, 34)	86	**93.4**
Ridge + Gabor [42]	*c27/rest all*	*90.9*	***93.9***
PLS-Gabor [26]	*c27/rest all*	*89.05*	***93.9***
3DMM + LGBP* [5]	*c27/c (11, 29, 07, 09, 05, 37)*	*99.0*	***100.0***



(a) Fiducials (b) Schematic pose estimation

Fig. 3. Schematic diagram to estimate the pose of a non-frontal face using fiducials.

The distance between the two eyes in frontal pose will be denoted by ee_1 and the distance between the nose and the lip by nl_1; similarly for the non-frontal pose to be estimated, the distance between the eyes is given by ee_2 and that between the nose and lip by nl_2. Assuming that the eyes, nose and lip are coplanar, i.e. the effect due to the nose sticking out is negligible, the new pose θ can be calculated as: $\theta = \arccos\left(\frac{ee_2/nl_2}{ee_1/nl_1}\right)$. A pictorial demonstration of this calculation is shown in Fig. 3.

To measure the poses in FERET and MultiPIE, manually annotated images were used to obtain the fiducial points and the frontal pose was used to calculate the rest of the non-frontal poses as explained above. The box and whisker plots for the estimated pose vs. the ground-truth pose for FERET and MultiPIE are shown in Figs. 4 and 5, respectively. It is clear that, in both databases, there are inconsistencies between the different subjects at the same pose, rendering both ground truth data inaccurate. The pose errors are higher in magnitude and scatter in FERET which is obtained under unconstrained conditions as compared to MultiPIE.

4.6. Pose estimation tolerance

Human head pose could be estimated in various ways besides using fiducial locations. However, it is necessary to get a sense of robustness and accuracy of the approach for a reliable estimate. Therefore, we empirically estimate the sensitivity of fiducial-based-pose-estimation scheme. The accuracy of the estimated pose depends on the accuracy with which the fiducial points are located. Therefore, it is necessary to estimate the induced error in the estimated pose due to the errors in the fiducial points location. It is done by randomly perturbing all four fiducial locations and re-estimating the pose using the perturbed fiducial locations. The error is defined as the absolute difference between the perturbed and originally estimated pose. The amount of perturbation for the eyes is a randomly chosen value between $\pm(x \times ee)$, i.e. fraction of the distance between the two eyes (ee). Similarly, nose and lips are perturbed by a randomly chosen value between $\pm(x \times nl)$, i.e. the same fraction of distance between the nose and lips (nl). The variation of average error over all the subjects and poses with increasing amount of perturbation fraction is shown in Fig. 6. We can see that the error in pose estimation is increasing with the increment in the fiducial location error but it is not very high and only after an error of 15% in fiducial locations, the pose estimation is severely affected.

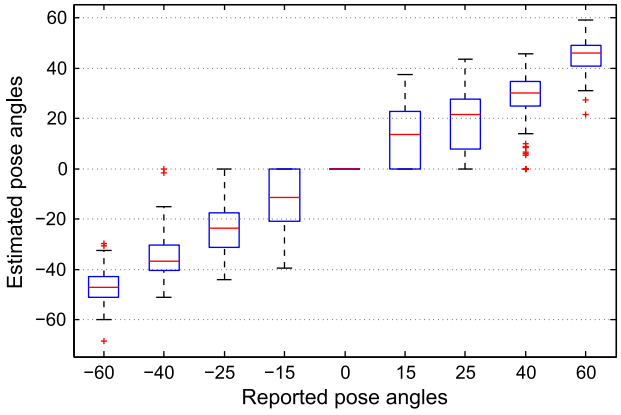


Fig. 4. Box and Whisker plot for pose errors on FERET data for all the nine poses.

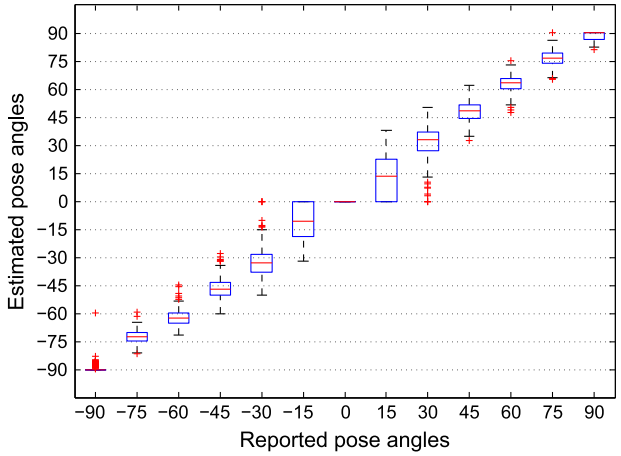


Fig. 5. Box and Whisker plot for pose errors on MultiPIE data for all the 13 poses which have only pitch variation from frontal.

5. Two-stage discriminative correspondence latent subspace

A discriminative representation approach such as LDA, requires multiple images per sample to learn the discriminative directions.

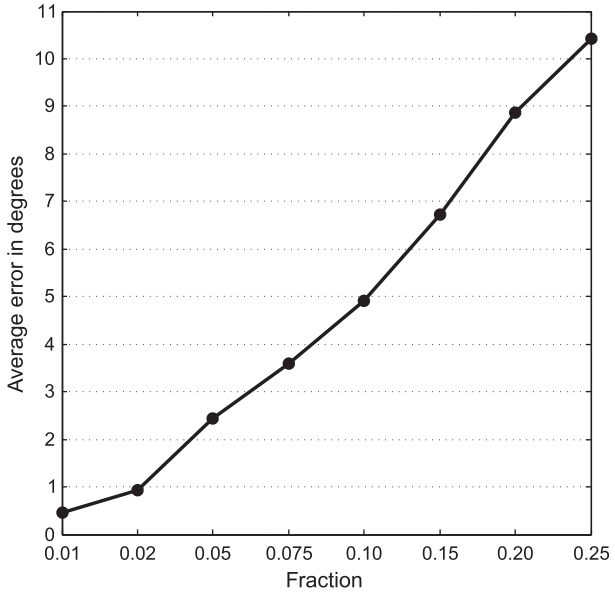


Fig. 6. Variation of pose estimation error with the amount of random perturbation in the fiducial locations.

We have a training set containing multiple images of a person but all the images are in different poses. Due to the loss of feature correspondence, we cannot use these multi-pose images directly to learn LDA directions. Results in [12] show that directly using them will lead to poor performance. However, we can learn a CLS for more than two poses simultaneously such that the projections of different pose images in the latent space have correspondence. Now, the multiple latent projections of a person can be used with LDA. Fortunately, using CCA as in (3), we can learn projectors for multiple poses to get a common CLS for a set of multiple poses. We empirically found that just using judiciously chosen set of poses (without LDA in latent space) to learn projectors offers some improvement over using only two poses. We defer the detailed discussion of selection of pose-sets and use of LDA to later sections. The multiple pose approach without LDA in latent space is termed Multiple CLS or MCLS and with LDA is termed Discriminative MCLS or DMCLS. The latent space projection $\mathbf{x}_l^i$ of ith subject in pose p $\left(\mathbf{x}_p^i\right)$ is given as

$$\mathbf{x}_l^i = W_p^T \mathbf{x}_p^i \tag{15}$$

Here, W_p^T is the projector for pose p and the subscript l indicates that $\mathbf{x}_l^i$ is in latent space. The projections of images in pose p using a projector for pose p are termed *same pose projections*. The latent space LDA offers discrimination based on the identity which is shown to be effective for classification [10,7].

The performance drop study also suggests that pose error is an important factor and needs to be handled for better performance. To tackle the pose error, we draw motivation from [9,47,8] where it has been shown that the inclusion of expected variations (those present in the testing set) in the training set improves the performance. Specifically, [9] has shown that using frontal and 30° training images with LDA improves the performance for 15° testing images. And, [8] shows that using artificially misaligned images, created by small random perturbation of fiducial points in frontal pose, during training with LDA offers robustness to small errors in fiducial estimation. We combine the two approaches and artificially simulate pose errors. Unfortunately, creating small pose errors is not as simple as creating fiducial misalignment in frontal images. We do it by deliberately projecting face images onto adjacent pose projectors to obtain *adjacent pose projections*. The dataset used has pose angle increments in steps of 15°; therefore, projection of a 45° image onto 30° and 60° projectors will give adjacent pose projections for 45°. The set of adjacent projections is given by

$$\mathcal{X}_l^i = \left\{\tilde{\mathbf{x}}_l^i : \tilde{\mathbf{x}}_l^i = W_{q\in A(p)}^T \mathbf{x}_p^i\right\} \tag{16}$$

here, $A(p)$ is the set of adjacent poses for pose p. The use of adjacent pose projections with LDA is expected to offer some robustness to small pose errors.

Same and adjacent pose projections have complementary information and both are important for robust pose-invariant face recognition. Therefore, we use both of them together as training samples with LDA to learn a discriminative classifier in the latent space. We call the resulting framework: Adjacent DMCLS of ADMCLS. ADMCLS is expected to offer robustness to pose errors smaller than 15° which is indeed the general range of pose errors observed in real-life as well as controlled scenarios. Apart from providing robustness to pose error, adjacent projection also provides more samples per class for better estimation of class mean and covariance. We empirically found that inclusion of pose error projections dramatically improves the performance on FERET and MultiPIE which is in accordance with [8] and our intuition. It also supports our claim that performance drop is due to pose errors. The complete flow diagram for the ADMCLS framework is depicted in Fig. 8.

5.1. Hyperparameter exploration

The proposed ADMCLS framework consists of two stages. The first stage involves learning the CLS and the second stage is learning the LDA directions using the projections in the latent subspace. Both stages have several different parameters, which will lead to different overall frameworks. For the ease of understanding and readability we summarize the names of different frameworks in Table 3. In this subsection we discuss the parameters involved and their effect on overall performance. We also discuss various

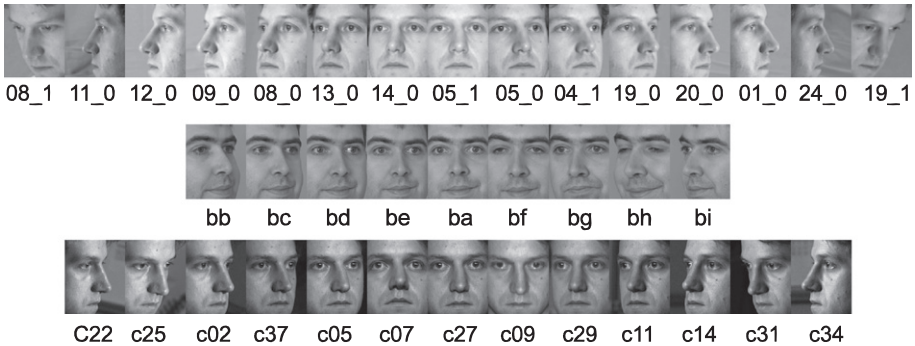


Fig. 7. Images with pose names, **MultiPIE** (top row), **FERET** (middle row) and**CMU PIE** (bottom row).

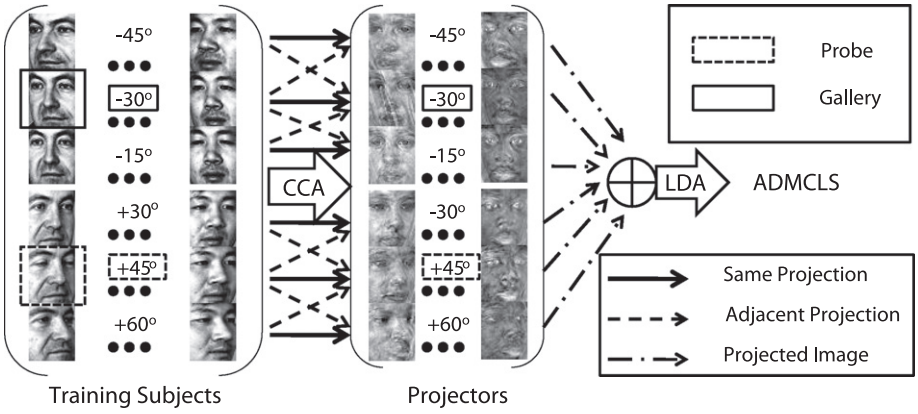


Fig. 8. The flow diagram showing the complete ADMCLS process pictorially for a pair of gallery (−30°) and probe (+45°) pose pair. The gallery and probe along with adjacent poses constitute the set of poses for learning the CLS (±30°, ±45°, −15°, and +60° for this case). Once the CLS is learned, same and adjacent pose projections (indicated by different arrow type) are carried out to obtain projected images in the latent subspace. An arrow from pose p images to pose q projector means projection of pose p images on pose q projector. All the projected images of a particular subject are used as samples in latent space LDA.

criteria to choose these parameters and their effect on the final performance.

To study the effect of a parameter, all the others were kept fixed but the one under study. Then the best values of individual parameters are used in the final framework. The final accuracy of the system in terms of rank-1 face identification rate is used as the performance measure to obtain the best value of each parameter. In order to facilitate future comparison of our approach, we have fixed the training subjects to be subject ID 1–34 for CMU PIE, 1–100 for MultiPIE and 1–100 (when arranged according to name) for FERET and made available the manually annotated fiducial points for FERET and MultiPIE used in our experiments. Testing is done on the rest of the subjects, i.e. 34, 237 and 100 testing subjects for CMU PIE, MultiPIE and FERET respectively.

5.1.1. Latent subspace dimension and learning model

The subspace dimension is an important parameter in all the subspace based methods and plays a critical role in performance. Too many dimensions can lead to over-fitting and too few to under-fitting; therefore, this parameter needs to be decided very carefully. There are some techniques based on the spectral energy of the eigen-system that can guide the proper selection such as – choosing a pre-defined ratio of energy to be preserved in the selected number of dimensions – rejecting the directions with lower eigen-value than a threshold. In the case of CCA, we selected top k eigen-vectors. We will see later that our final framework does not require a very careful selection of this parameter and is pretty robust to its variation. In the case of PLS we are using an iterative greedy algorithm and the number of dimensions can be selected by using only those directions which contain some pre-specified amount of total variation. However, it was observed that beyond a certain number of dimensions the accuracy remains constant. For BLM, we can use the spectral energy approach to select the number of dimensions. The selected number of dimensions of the CLS would be indicated as a superscript of the final framework name.

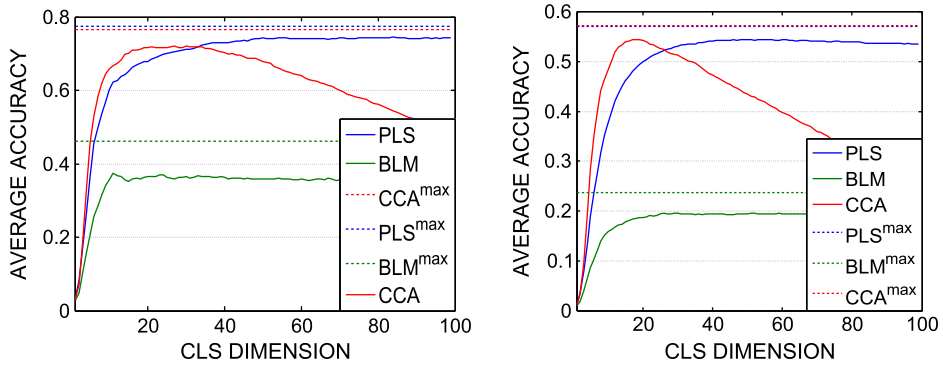
To keep things simple we have used two poses and 1-NN matching as the constituents of the final framework and varied the number of dimensions of CLS. The accuracy is the average accuracy for all possible gallery-probe pairs for the same number of CLS dimensions. There are 15 poses in MultiPIE so there is a total of 210 gallery-probe pose pairs and 72 for FERET (nine poses). The variation of accuracy for PLS, CCA and BLM on FERET and MultiPIE is shown in the Fig. 9a and b. It is obvious that different gallery-probe pairs will achieve the maximum accuracy with different number of CLS dimensions but we are calculating the average accuracy by considering the same CLS dimension for all pairs. To show the difference between our performance measure and the best possible accuracy obtained by using different CLS dimensions for different gallery-probe pairs, we calculated the best accuracy for all the pose pairs and averaged them to get the overall accuracy. These best accuracies are plotted as dashed horizontal lines in the same figure.

The choice of learning model has significant impact on the overall performance. We investigated three different choices for learning method: CCA, PLS and BLM and found that PLS performed slightly better than CCA for pose invariant face recognition and BLM is the worst performing [11]. However, PLS cannot be used to learn a CLS framework for more than two poses which makes it useless for the MCLS framework and BLM performs significantly worse than CCA. So, we used CCA for the cases when more than two poses are used for training.

Fig. 9 clearly reveals the effect of learning model on face identification rate. The most important and satisfying observation is that the maximum possible accuracy is not significantly higher than the average accuracy justifying our assumption of equal CLS dimension across all gallery/probe pose pairs. Clearly, BLM performance is significantly worse than CCA and PLS which is in accordance with the results obtained in [11]. The performance of CCA and PLS is al-

Table 3
Framework names based on the components used, the super-script in the name denotes the CLS dimension.

Name	Model	Training set poses	Projections	Classifier	CLS dimension
CCA^{10}	CCA	Gallery + probe	Same pose	1-NN	10
PLS^{10}	PLS	Gallery + probe	Same pose	1-NN	10
BLM^{20}	BLM	Gallery + probe	Same pose	1-NN	20
$MCLS^{10}$	CCA	Gallery + probe + intermediate	Same pose	1-NN	10
$DMCLS^{40}$	CCA	All poses	Same + adjacent pose	LDA	40
$ADMCLS^{10}$	CCA	Gallery + probe + adjacent	Same + adjacent pose	LDA	10



(a) FERET (b) Multi PIE

Fig. 9. Result of CLS based recognition using 1-NN classifier on FERET and MultiPIE. (CCA/PLS/BLM)^{max} represents the maximum possible accuracy using different number of CLS dimensions for all gallery-probe pairs. For MultiPIE, PLS^{max} and CCA^{max} overlap and only one of them is visible.

most similar for MutliPIE and PLS performs better than CCA for FERET which is also in accordance with [11]. One clear observation from the figure is that CCA performance is sensitive to CLS dimension and achieves maxima in a short range. On the other hand, the performance of BLM and PLS increase till a certain number of dimensions and then stays nearly constant. This brings out the fact that CCA is prone to over fitting while BLM and PLS are not (see Fig. 9).

5.1.2. Set of training poses

This has some effect on the obtained projectors since different sets of training poses will generate somewhat different projectors for each pose pair. Moreover, the supervised classifier in the latent space uses the projections as samples hence, it will have some bearing on the classifier too. In the case of PLS as the learning model, we can have only two training poses because of poor learning for multiple poses but this is not a problem with BLM or CCA. The set of poses used for training has deep impact on the obtained CLS performance and further improvements. We indicate the use of multiple training poses in the framework by preceding CLS by *M*, i.e. MCLS. Fig. 10 visually brings out the existence of correspondence between the coupled subspaces using CCA.

The intuition of using more than two training poses can be understood in terms of robustness to noise offered by additional poses for CCA. It was pointed out and proved in [34] in a completely different context of clustering that adding more styles of data improves noise-robustness which also holds in our case of pose variation. As explained earlier in Section 3.2, CCA based CLS is a way of learning correspondence by maximizing correlation. The correlation between the training images in two different poses are most likely due to two factors: true correspondence and noise. We ideally want that the correlation is only due to correspondence. However, our data always contains some noise in the form of pose errors and/or inaccurate fiducial location. Presence of noise in the data can cause spurious correlations leading to false correspondence that will affect the performance. When more than two poses are used simultaneously, the obtained correlation between these poses has a higher probability of being due to correspondence because it is present in all the poses. However, this does not mean that we should add too many poses because it will decrease the flexibility of the learning model and lead to under-fitting. Thus, two poses will lead to over-fitting and too many will cause under-fitting, hence we choose four poses to strike a balance. Note that, the value four came out of empirical observation.

To evaluate the effect of changing the sets of training poses on the final framework for a particular gallery-probe pair, we include poses other than gallery and probe poses to learn CLS. This procedure raises some interesting questions: which poses should be included in training set? how many poses should be used? To answer these questions, we adopt a very simple approach that illustrates the effect of using multiple training poses. We use three gallery poses and all the possible probe poses for the selected gallery poses. For FERET, we choose pose **ba**(frontal), **bd** (25°) and **bb** (60°) and for MultiPIE, we choose 051(frontal), 190(45°) and 240(90°) as gallery poses. In addition to the gallery and probe we also select adjacent intermediate poses based on the viewing angle, i.e. if we have gallery as frontal (0°) and probe as +60° then we take two additional poses to be +15° and +45°. Similarly, for gallery as frontal and probe as +30° we take only one additional pose +15° since it is the only intermediate pose.

Once the latent subspace is learned we use 1-NN for classification. The number of CLS dimensions is kept at 17 so the final frameworks are called as MCLS^{17}. We show the comparison of CCA based MCLS^{17} vs. CCA^{20} in Fig. 11a and b for FERET and MultiPIE respectively. There are some missing points in the performance curves in both figures because an adjacent gallery-probe pose pair does not have any intermediate pose. The comparison clearly highlights the improvement offered by using multiple poses for learning the latent subspace. We generally observe some improvement with MCLS^{17} framework for gallery and probe poses with large pose difference except for few places where it either remained the same or decreased slightly. We also observe that the improvement is more significant in FERET as compared to MultiPIE which is due to the fact that MulitPIE dataset has less pose errors than FERET, as shown in Section 4.5. Therefore, MCLS framework has more to offer in terms of robustness to pose errors in FERET as compared to MultiPIE.

The second stage of the framework is learning a supervised classifier using the latent subspace projections. This stage has two crucial parameters: Set of projections and Classifier. The next two sections explore their affect on the performance.

5.1.3. Set of projections and classifier

It refers to the combination of the set of latent subspace projections for a subject and the classifier used for matching. As discussed earlier, we have two choices for projecting a face image in the CLS and both contain complementary information which can be utilized by a classifier for recognition. Since all the databases used in this paper have pose angles quantized in steps of 15°, the difference between any two adjacent poses is 15°. In our framework, we do not consider more than 15° pose difference because they will render the projection meaningless and they do not exist in real life scenarios.

Fig. 10. Projector bases corresponding to top eigen-values obtained using CCA (first five rows) and PCA [32] (bottom five rows) obtained using 100 subjects from FERET. CCA projectors are learned using all the poses simultaneously and PCA projectors are learned separately for each pose. Each row shows the projector bases of the pose for equally indexed eigen-value. Observe that, projector bases are hallucinated face images in different poses and the CCA projector bases look like rotated versions of the same hallucinated face but there is considerable difference between PCA projectors. This picture visually explains the presence of correlation in the latent CLS space using CCA and its absence using PCA.

As mentioned earlier, CCA is used as the learning model for all the experiments with more than two poses in the training. MultiPIE has 15 poses and FERET has nine, so the size of the eigen-system for MultiPIE becomes too big and requires large memory. So, all the exploratory experiments were done with FERET and conclusions were used to decide the optimal strategy for MultiPIE. In order to avoid under-fitting we adopt a simple strategy to select a subset of poses for training that is based on gallery-probe pair. The gallery-probe pairs along with the adjacent poses of them are selected as the training set of poses. So, for a $+45°/-30°$ gallery/probe pair the training set would be $\pm30°$, $\pm45°$, $+60°$, $-15°$ and for $-15°/0°$ training pose set is $\pm15°$, $0°$, $30°$. Adjacent poses are selected to simulate pose error scenario. We call this variant of DMCLS as Adjacent Discriminant Multiple Coupled Subspace (ADMCLS). To evaluate the effect of different latent space projections, we plot the average accuracy across all 72 gallery/probe pairs in Fig. 12 for the following settings: 1-NN classifier with two poses denoted by CLS; Intermediate poses and 1-NN classifier denoted by MCLS; two poses and LDA denoted by DCLS; all nine poses for FERET and adjacent projections with LDA denoted by DMCLS and adjacent set of training poses with adjacent projections and LDA denoted by ADMCLS.

It is clear from the Fig. 12 that ADMCLS performs the best closely followed by DMCLS, while, CLS is the worst performing approach with DCLS and MCLS performance being slightly better than CLS. The use of LDA with adjacent projections did not only increase the accuracy significantly but also makes the final framework fairly insensitive to CLS dimension, which eliminates the burden of determining it by cross-validation. This significant improvement is due to artificial simulation of pose error scenarios and learning to effectively neglect such misalignments for classification using LDA. One more reason contributing to the improvement is the LDA assumption of similar within-class covariance for all the classes. In our case, indeed the within-class covariance matrices are almost the same because the samples of all the classes in CLS are obtained using same set of CLS bases and the types of projection are also the same for all the classes. The recognition rates for all

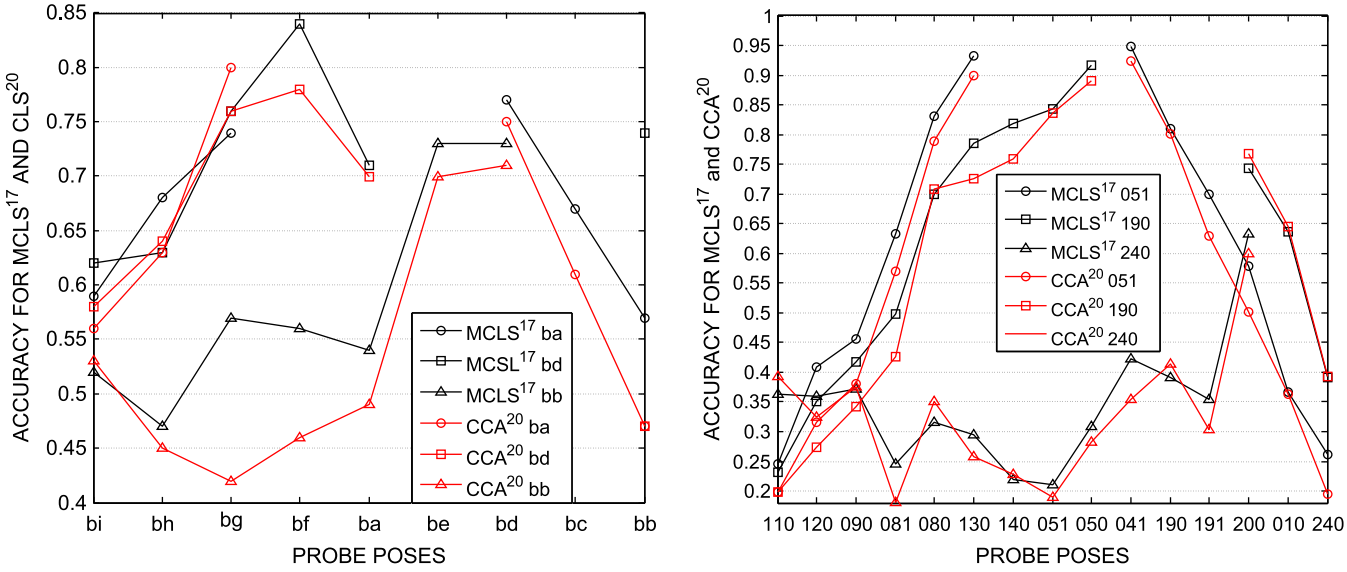


Fig. 11. Comparison of $MCLS^{17}$ vs. CCA^{20} with varying gallery-probe pairs for (a) three gallery poses ba (frontal), bd (40°) and bb (60°) on FERET dataset. (b) Three gallery poses 051(frontal), 190(45°) and 240(90°) on MultiPIE dataset. $MCLS^{17}ba$ indicates that the gallery is pose **ba**, **multiple** poses are used during training and CCA is the learning model with **17** dimensional CLS and **1-NN** classifier while $CCA^{20}ba$ indicates that the gallery is pose **ba**, **two** poses are used during training and **CCA** is the learning model with **18** dimensional CLS and **1-NN** classifier

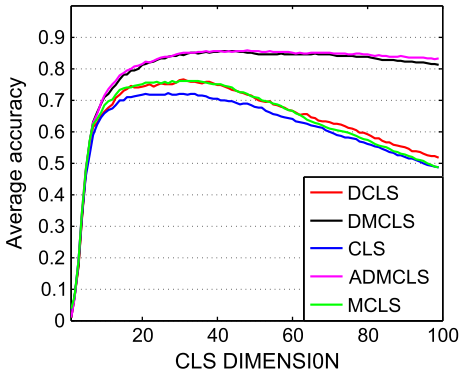


Fig. 12. Variation of CLS, MCLS, DCLS, DMCLS, and ADMCLS accuracy with latent space dimension for all the gallery-probe pairs on FERET.

the 72 pose pairs with $DMCLS^{40}$ using all the pose pairs in training set are given in Table 4. To prove the point that the improvement is actually due to handling pose errors we also obtain the relative improvement by $ADMCLS^{40}$ over CLS^{22} for all gallery-probe pairs. The difference is plotted as a heat map for better visualization in Fig. 13a. From the figure, it is evident that the most significant improvements are in the cases where either the gallery or the probe pose is far away from frontal pose. In these cases, the chance and extent of pose errors and incorrect fiducial locations is most likely and prominent (see Table 5).

5.2. Computational complexity

It is obvious that learning an ADMCLS with multiple poses offers various advantages but it also requires some additional computational cost. The computational bottleneck of the ADMCLS framework is the solution of the generalized eigen-value problem in (3). The complete generalized eigen-value decomposition of a pair of $N \times N$ square matrices (A,B) is $O(N^3)$ but we only need the leading k eigen-vectors. Therefore, the cost comes down to $O(kN^2)$. In our case, $N = \sum_m D_m$ where, D_m is the dimension of the mth pose feature space (number of pixels in our experiments). For simplicity, let us assume that the dimension of each pose feature space is equal to a constant D. Therefore, $N = MD$, where M is the number of coupled poses. Hence, the computational complexity as a function of the number of coupled poses M and the dimension of feature space is $O(kD^2M^2)$.

6. Experimental analysis

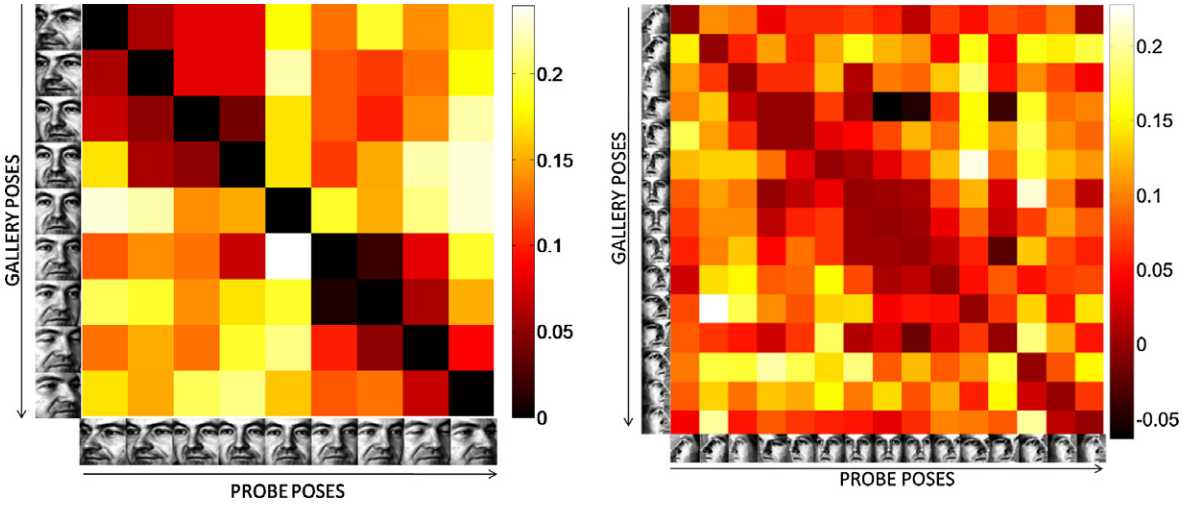
In this section we provide the rank-1 indentification rates obtained on CMU PIE, FERET and MultiPIE using best parameters settings and compare our results with prior work on the same datasets. Please note that, CCA is used as the learning model for all the methods using more than two poses in training set, for the reasons explained in previous sections.

6.1. Training and testing protocol

Like any other learning based approach we require training data to learn the model parameters. We assume access to a training data that has multiple images of a person under different poses and ground-truth poses of training as well as testing faces. Although fiducial points can be used for a better estimation of pose, we use the ground-truth poses for a fair comparison with previous approaches. Moreover, automatic pose estimation algorithms and fiducial detectors always have some error. Therefore, working with small pose errors reflects performance with automatic pose or fiducial detector. CMU PIE, FERET and MultiPIE have multiple images of a person under a fixed set of poses. Hence, we use some part of the data as training and the rest as testing. We also need to align the faces under different poses which requires fiducial landmark points. In the training phase, we obtain the projectors for all the possible gallery/probe pose pairs for the required framework, i.e. ADMCLS, DMCLS, etc. At testing time, we assume that the gallery and probe poses are known and use appropriate projectors for projection followed by matching. For testing purpose we always project the images on the same pose projector as per as the ground-truth poses. For a completely automatic face recognition

Table 4
DMCLS^{40}/*ADMCLS*^{40} for all possible gallery-probe pairs on FERET.

Pose Angle	bi −60°	bh −40°	bg −25°	bf −10°	ba 0°	be 10°	bd 25°	bc 40°	bb 60°	DMCLS^{40}Avg/ ADMCLS^{40} Avg
bi	−/−	98/98	92/93	88/82	70/77	81/80	79/80	76/69	70/63	**81.75**/80.25
bh	97/97	−/−	99/99	94/94	80/84	90/87	79/77	71/70	62/60	**84.00**/83.50
bg	95/96	97/99	−/−	100/100	91/92	98/97	90/92	78/76	68/68	89.63/**90.00**
bf	83/91	93/95	96/99	−/−	93/97	97/99	95/95	85/84	73/71	89.38/**91.37**
ba	75/79	77/85	89/94	91/96	−/−	90/95	87/94	81/82	67/70	82.13/**86.38**
be	86/83	91/88	96/96	98/99	90/99	−/−	99/100	97	84	92.50/**93.25**
bd	79/78	84/83	90/90	91/95	90/89	98/98	−/−	98	84/86	89.25/**89.63**
bc	75/70	73/67	77/73	82/79	80/80	92/94	97/97	−/−	95/96	**83.88**/82.00
bb	71/70	66/60	67/62	67/67	64/65	81/82	82/84	95/95	−/−	**74.13**/73.12



(a) FERET improvement map (b) MultiPIE improvement map

Fig. 13. Improvement map for (a) using ADMCLS^{40} over CCA^{20} for FERET and (b) using ADMCLS^{25} over CCA^{18} for MultiPIE. The original accuracies were all between 0 (0%) and 1 (100%). It is evident from the two maps that the amount of improvement is more in FERET as compared to MultiPIE. Also, the improvement is more when either the gallery or probe pose is far from the frontal view.

Table 5
Comparison of ADMCLS^{40} with other published works on feret with frontal gallery.

Method	Probe pose								
	bi	bh	bg	bf	be	bd	bc	bb	Avg
LDA [13]	18.0	55.0	78.0	95.0	90.0	78.0	48.0	24.0	60.8
LLR [13]	45.0	55.0	90.0	93.0	90.0	80.0	54.0	38.0	68.1
CCA [13]	65.0	81.0	93.0	94.0	93.0	89.0	80.0	65.0	82.5
Stack [43]	40.0	67.5	88.5	**96.5**	94.5	86.0	62.5	38.0	71.7
Yamada [25]	8.5	32.5	74.0	88.0	83.0	54.0	23.5	6.5	46.3
Ridge + Int [42]	67.0	77.0	90.0	91.0	92.0	89.0	78.0	69.0	81.6
DMCLS^{40}	75.0	77.0	89.0	91.0	90.0	87.0	81.0	67.0	82.1
ADMCLS^{40}	**79.0**	**85.0**	**94.0**	96.0	**95.0**	**90.0**	**82.0**	**70.0**	**86.4**
3DMM [4]	*90.7*	*95.4*	*96.4*	*97.4*	*99.5*	*96.9*	*95.4*	*94.8*	*95.8*
Ridge + Gab [42]	*87.0*	*96.0*	*99.0*	*98.0*	*96.0*	*96.0*	*91.0*	*78.0*	*92.6*
3DMM-LGBP [5]	–	*90.5*	*98.0*	*98.5*	*97.5*	*97.0*	*91.9*	–	*95.6*

system, pose and fiducial landmarks should be obtained automatically. However, for experimentation purposes, we assume them to be known beforehand, a common practice followed in much previous work [15,16,42,25,37,43,14,13,9,26]. Fortunately, research and commercial systems have shown impressive performance in automatic pose and fiducial determination that can be used in conjunction with our approach to make an automatic pose invariant face recognition system.

6.2. FERET

This dataset contains 200 subjects in nine different poses spanning ±60° view-point. All the images for one person along with the pose name are shown in Fig. 7. Pre-processing steps similar to CMU PIE were used except that the final facial region crops are of size 50×40 pixels. Subjects 1–100 were chosen as training subjects and 101–200 as testing. Since, there are nine poses, we have 72 different gallery-probe pairs.

We report the accuracy for FERET data set using two different variants of DMCLS to bring out the fact that using more than the required number of poses in training may lead to poor performance. We report DMCLS based accuracy which uses all the nine poses in the training and adjacent projection based LDA in latent space and ADMCLS based accuracy which uses a subset of poses for training. The number of CLS dimension is indicated as the superscript and CCA is used as the learning model. Table 4 reports the accuracy

Table 6
MultiPIE accuracy for all possible 210 gallery-probe pairs using ADMCLS[25] with 237 testing subjects. The duplet below the pose name indicates the horizontal, vertical angle, i.e. 45, 15 means 45° horizontal and 15° vertical angle.

Prb→ Gal↓	110 −90, 0	120 −75, 0	090 −60, 0	081 −45, 45	080 −45, 0	130 −30, 0	140 −15, 0	051 0, 0	050 15, 0	041 30, 0	190 45, 0	191 45, 45	200 60, 0	010 75, 0	240 90, 0	**Avg**
110	−/−	76.4	65.8	34.6	48.5	37.6	33.3	27.4	21.9	31.6	31.2	24.9	35.9	49.4	43.9	**37.5**
120	78.5	−/−	81.9	48.5	68.8	57.8	54.9	43.9	42.2	44.7	44.7	27.4	59.1	65.0	50.2	**51.2**
090	67.1	81.9	−/−	59.5	80.2	72.2	51.9	46.0	46.8	54.0	55.3	32.1	64.1	60.8	43.0	**54.3**
081	38.0	49.8	57.8	−/−	78.5	82.3	73.8	55.7	48.9	52.3	57.0	63.7	49.8	40.1	28.7	**51.8**
080	55.3	70.9	78.9	76.8	−/−	97.9	93.2	85.7	84.8	82.7	84.0	54.0	72.6	59.9	40.1	**69.1**
130	39.7	58.6	72.6	84.4	97.0	−/−	96.2	93.7	92.8	90.7	86.9	60.8	68.4	54.9	33.8	**68.7**
140	30.4	52.7	57.0	73.8	90.7	97.5	−/−	98.7	95.4	92.8	89.0	60.8	64.1	45.6	24.1	**64.8**
051	27.0	42.2	48.5	58.6	84.8	96.6	99.2	−/−	99.2	96.2	89.0	65.0	57.4	47.7	27.8	**62.6**
050	25.7	40.9	47.7	54.0	85.2	95.4	97.5	98.7	−/−	98.7	94.9	74.7	75.1	59.5	35.9	**65.6**
041	26.6	50.2	51.9	52.3	81.0	93.7	95.8	94.9	98.7	−/−	96.6	88.6	80.6	72.6	43.9	**68.5**
190	27.4	50.2	51.9	53.2	78.9	86.1	89.9	87.8	94.5	97.5	−/−	85.7	90.3	70.0	53.6	**67.8**
191	22.8	30.8	30.8	65.0	49.8	65.8	60.8	62.4	70.0	87.3	83.1	−/−	77.2	63.3	39.2	**53.9**
200	36.3	59.1	65.8	52.3	72.2	67.9	63.7	58.6	72.2	84.4	87.3	81.0	−/−	97.0	75.1	**64.9**
010	44.7	63.7	61.6	43.0	64.6	53.2	47.7	54.0	63.7	77.6	75.5	65.4	95.4	−/−	94.9	**60.3**
240	43.5	52.3	43.0	26.6	41.8	31.6	28.3	22.4	34.6	45.6	51.1	38.8	79.7	93.2	−/−	**42.2**

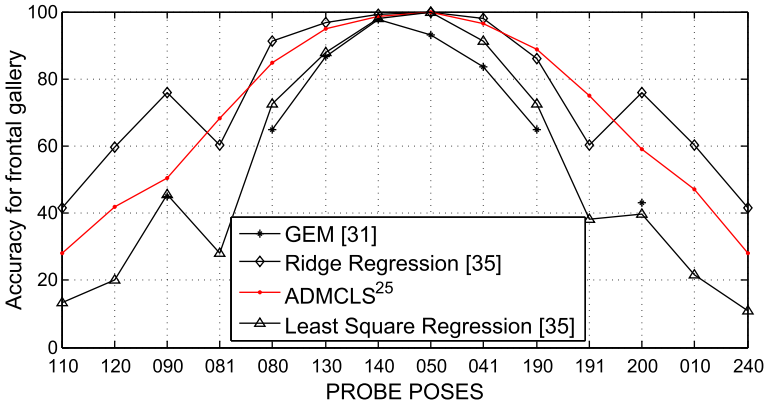


Fig. 14. Comparison of *ADMCLS*[25] with other approaches on MultiPIE dataset with frontal gallery.

for all possible gallery-probe pairs using the two different variant, i.e. DMCLS and ADMCLS. The table clearly indicates the advantage of using ADMCLS over DMCLS when near frontal poses are used as gallery pose. It also indicates that when extreme poses are gallery then using DMCLS is slightly better than ADMCLS, a possible explanation is that extreme poses require more regularization than flexibility. Table 5 reports the comparison between the proposed approach and past approaches for pose invariant face recognition on FERET. We report the accuracy obtained using 3DMM [4] approach to indicate the performance difference between 2D and 3D approaches. The difference in performance between 2D and 3D approaches supports the fact that 3D information improves performance in pose invariant face recognition.

The results of [42] are shown under two settings: with and without Gabor features. The authors have extracted Gabor features at five hand annotated fiducial locations using five scales and eight orientations resulting in 200 local classifiers which they fuse using the technique given in [25]. The method involves modeling the conditional probability of the Gabor response g_i of classifier i for same and different identities, i.e. $P(g_i|same)$ and $P(g_i|dif)$ respectively. Then, Bayes Rule is used to obtain posteriors $P(same|g_i)$ and $P(dif|g_i)$ and the probability of final classification is the sum of the posterior probabilities. The inclusion of Gabor features has improved the accuracy dramatically because they are more discriminative than intensity features. Moreover, using Gabor features at hand-annotated fiducial landmarks is providing manual correspondence to the learning method. Combining Gabor features with probabilistic fusion is interesting and worth trying within our framework. Surprisingly, for CMU PIE our simple PLS based approach even outperformed the Gabor feature based approach.

6.3. Multi PIE

MultiPIE is an extension of CMU PIE data set containing more subjects and more pose-variation. It has a total 337 subjects photographed in four different sessions, under 15 different poses, 20 illumination conditions and four different expressions. We only took neutral expression and frontal lighting images for our experiments. All the pre-processing steps are the same as in CMU PIE except that the cropped facial region is 40×40 pixels. We took subject ID 1–100 as training and 101 to 346 as testing, resulting in a total of 237 testing subjects. For MultiPIE we could not obtain MCLS using all the poses in the training set due to memory problem associated with large eigen-value problem. Hence, we adopt the ADMCLS approach to select a subset of training poses and report the accuracy in Table 6. The MultiPIE data is relatively new and not many results are reported for pose invariant face recognition on it. We show our results along with the results of other works in Fig. 14. It should be noted that we are reporting the results of [42] with pixels intensities as feature.

Interestingly, our 2D approach is better than the 3D GEM [38] approach. We also observe that our approach is comparable to the approach in [42] for small pose differences but the difference increases with the pose angle. This might be due to the fact they report their result under frontal gallery and non-frontal probe only, giving them the opportunity to better tune the parameter but we report the results under general pose variation and do not optimize

our method for frontal gallery and non-frontal pose. Moreover, we have outperformed [42] on both CMU PIE and FERET by large margins without optimizing for the case of frontal gallery images.

7. Conclusion and discussion

We have proposed a generic Discriminative Coupled Latent Subspace based method for pose invariant face recognition. The learned set of coupled subspaces projects the images of the same person under different poses to close locations in the latent space, making recognition possible using a simple 1-NN or discriminative learning. We have discussed the conditions for such projection directions to exist and perform accurately. We further exploit the property of CCA to couple more than two subspaces corresponding to different poses and show that judiciously using multiple poses to learn the coupled subspace performs better than using just two poses. That is because information from multiple views is more consistent and robust to noise (pose errors and incorrect fiducials) than just two views. Multiple coupled subspaces also provide us with the opportunity to generate multiple samples of a person in the latent subspace which can be used with LDA to encode discriminative information.

We have provided empirical evidence that pose-invariant-face recognition suffers from pose errors even under controlled settings, leading to poor performance. We tackle the pose error problem by artificially simulating pose error scenarios via adjacent-pose-latent projection. The latent projections obtained by projecting the images of a person under different poses on the same and adjacent pose projectors are used with LDA to effectively avoid the drop in performance due to small pose errors. The proposed approach has achieved state-of-the-art results on CMU PIE and FERET when four fiducial points are used with simple intensity features and comparable results on MultiPIE.

We experiment with pose variation only and illumination is considered to be constant. However, owing to the independent block structure of the overall framework, it can be easily extended to handle lighting variations by using some illumination invariant representation such as: The Self Quotient Image [50] and Oriented gradient [51]. Moreover, Gabor features extracted at specific fiducial locations can be used to improve the performance further as in [42,15,16,26,5]. The coupled subspaces are learned in generative manner and only after projection on these subspaces, label information is used with LDA. The method could be improved by learning a discriminative coupled subspace directly. Learning such a subspace and using it for pose and lighting invariant face recognition is one of our future endeavors.

References

[1] P. Santemiz, L.J. Spreeuwers, N.J.R. Veldhuis, Side-view face recognition, in: Proceedings of 32nd WIC Symposium on Information Theory in the Benelux, 10–11 May 2011.

[2] X. Jhang, Y. Gao, Face Recognition Across Pose: A Review, Pattern Recognition, vol. 2, Elsevier, 2009. pp. 2876–2896.

[3] A. Wagner, J. Wright, A. Ganesh, Z. Zhou, Y. Ma, Towards a practical face recognition system: Robust registration and illumination by sparse representation, in: Proceedings of IEEE CVPR, 2009, pp. 597604.

[4] V. Blanz, T. Vetter, Face recognition based on fitting a 3d morphable model, IEEE Trans. Patt. Anal. Mach. Intel. 25 (9) (2003) 1063–1074.

[5] A. Asthana, T.K. Marks, M.J. Jones, K.H. Tieu, Rohith MV, Fully automatic pose-invariant face recognition via 3D pose normalization, in: Proceedings of IEEE ICCV, 2011, pp. 937–944.

[6] The Facial Recognition Technology (FERET) Database <http://www.itl.nist.gov/iad/humanid/feret/feret_master.html>.

[7] D.L. Swets, J. Weng, Using discriminant eigen features for image retrieval, IEEE Trans. Patt. Anal. Mach. Intel. 18 (8) (1996) 831–836.

[8] S. Shan, Y. Chang, W. Gao, B. Cao, P. Yang, Curse of mis-alignment in face recognition: problem and a novel mis-alignment learning solution, in: IEEE Conference on Auto Face Gesture Recognition, 2004, pp. 314–320.

[9] A. Sharma, A. Dubey, P. Tripathi, V. Kumar, Pose invariant virtual classifiers from single training image using novel hybrid-eigenfaces, Neurocomputing 73 (10) (2010) 1868–1880.

[10] P.N. Belhumeur, J. Hespanha, D.J. Kriegman, Eigenfaces vs. Fisherfaces: recognition using class specific linear projection, IEEE Trans. Patt. Anal. Mach. Intel. 19 (1997) 711–720.

[11] A. Sharma, D.W. Jacobs, Bypassing synthesis: PLS for face recognition with pose, low-resolution and sketch, in: Proceedings of IEEE CVPR, 2011, pp. 593–600.

[12] X. Chai, S. Shan, X. Chen, W. Gao, Locally linear regression for pose invariant face recognition, IEEE Trans. Image Process. 16 (7) (2007) 1716–1725.

[13] A. Li, S. Shan, X. Chen, W. Gao, Maximizing intra-individual correlations for face recognition across pose differences, in: Proceedings of IEEE CVPR, 2009, pp. 605–611.

[14] S. Lucey, T. Chen, A viewpoint invariant, sparsely registered, patch based, face verifier, Int. J. Comput. Vision 80 (2008) 58–71.

[15] S.J.D. Prince, J.H. Elder, J. Warrell, F.M. Felisberti, Tied factor analysis for face recognition across large pose differences, IEEE Trans. Patt. Anal. Mach. Intel. 30 (6) (2008) 970–984.

[16] S.J.D. Prince, P. Li, Y. Fu, U. Mohammed, J. Elder, Probabilistic models for inference about identity, IEEE Trans. Patt. Anal. Mach. Intel. 34 (1) (2012) 144–157.

[17] P. Phillips, H. Wechsler, J. Huang, P.J. Rauss, The FERET database and evaluation procedure for face recognition algorithms, Image Vision Comput. 16 (1998) 295–306.

[18] J.B. Tenenbaum, W.T. Freeman, Separating style and content with bilinear models, Neural Comp. 12 (6) (2000) 1247–1283.

[19] D.R. Hardoon, S.R. Szedmak, J. Shawe-Taylor, Canonical correlation analysis: an overview with application to learning methods, Neural Comput. 16 (2004) 2639–2664.

[20] R. Rosipal, N. Krämer, Overview and recent advances in partial least squares, in: Subspace, Latent Structure and Feature Selection Techniques, Lecture Notes in Computer Science, Springer, 2006. pp. 34–51.

[21] Partial Least Square Tutorial <http://www.statsoft.com/textbook/partial-least-squares/#NIPALS>.

[22] A.L. Boulesteix, K. Strimmer, Partial least squares: a versatile tool for the analysis of high-dimensional genomic data, Briefings Bioinform. 8 (1) (2006) 32–44 (Advance Access publication).

[23] J. Shawe-Taylor, N. Christianini, Kernel Methods for Pattern Analysis, Cambridge University Press, 2004.

[24] T. Sim, S. Baker, M. Bsat, The CMU pose, illumination, and expression database, IEEE Trans. Patt. Anal. Mach. Intel. 25 (12) (2003) 1615–1618.

[25] T. Kanade, A. Yamada, Multi-subregion based probabilistic approach toward pose-invariant face recognition, in: Proceedings of IEEE CIRA, 2003, pp. 954–959.

[26] A. Li, S. Shan, X. Chen, W. Gao, Cross-pose face recognition based on partial least squares, Pattern Recog. Lett. 32 (15) (2011) 1948–1955.

[27] C. Dhanjal, S.R. Gunn, J.S. Taylor, Efficient sparse kernel feature extraction based on partial least squares, IEEE Trans. Patt. Anal. Mach. Intel. 31 (8) (2009) 1947–1961.

[28] J. Baeka, M. Kimb, Face recognition using partial least squares components, Pattern Recog. 37 (2004) 1303–1306.

[29] V. Struc, N. Pavesic, Gabor-based kernel partial-least-squares discrimination features for face recognition, Informatica 20 (1) (2009).

[30] X. Li, J. Ma, S. Lia, Novel face recognition method based on a principal component analysis and kernel partial least square, IEEE ROBIO (2007) 1773–1777.

[31] W.R. Schwartz, H. Guo, L.S. Davis, A robust and scalable approach to face identification, in: Proceedings of ECCV, 2010, pp. 476–489.

[32] M. Turk, A. Pentland, Eigenfaces for recognition, J. Cog. Neurosci. 3 (1) (1991) 71–86.

[33] R. Gross, I. Matthews, J. Cohn, T. Kanade, S. Baker, MultiPIE, Image Vision Comput. 28 (5) (2010) 807–813.

[34] M.B. Blaschko, C.H. Lampert, Correlational spectral clustering, in: Proceedings of IEEE CVPR, 2008, pp. 1–8.

[35] C.D. Castillo, D.W. Jacobs, Using stereo matching with general Epipolar geometry for 2-D face recognition across pose, IEEE Trans. Patt. Anal. Mach. Intel. 31 (12) (2009) 2298–2304.

[36] C.D. Castillo, D.W. Jacobs, Wide-baseline stereo for face recognition with large pose variation, in: Proceedings og IEEE CVPR, 2011, pp. 537–544.

[37] R. Gross, I. Matthews, S. Baker, Appearance-based face recognition and light-fields, IEEE Trans. Patt. Anal. Mach. Intel. 26 (4) (2004) 449–465.

[38] U. Prabhu, J. Heo, M. Savvides, Unconstrained pose invariant face recognition using 3D generic elastic models, IEEE Trans. Patt. Anal. Mach. Intel. 33 (10) (2011) 1952–1961.

[39] T. Ojala, M. Pietikainen, T. Maenpaa, Multiresolution gray-scale and rotation invariant texture classification with local binary patterns, IEEE Trans. Patt. Anal. Mach. Intel. 24 (7) (2002).

[40] D. Lowe, Distinctive image features from scale-invariant keypoints, International Journal of Computer Vision 60 (2) (2004) 91–110.

[41] L. Wiskott, J. Fellous, N. Kruger, C. Von der Malsburg, Face recognition by elastic bunch graph matching, IEEE Trans. Patt. Anal. Mach. Intel. 19 (7) (1997).

[42] A. Li, S. Shan, W. Gao, Coupled bias-variance trade off for cross-pose face recognition, IEEE Trans. Image Process. 21 (1) (2012) 305–315.

[43] A.B. Ashraf, S. Lucey, T. Chen, Learning patch correspondences for improved viewpoint invariant face recognition, in: Proceedings of IEEE CVPR, 2008, pp. 1–8.
[44] Q. Ying, X. Tang, J. Sun, An associate-predict model for face recognition, in: Proceedings of IEEE CVPR, 2011, pp. 497–504.
[45] S. Baker, I. Matthews, Lucas-kanade 20 years on: a unifying framework, Int. J. Comput. Vision 56 (3) (2004) 221–255.
[46] C.M. Bishop, Pattern recognition and machine learning, first ed., Springer, 2006.
[47] D.L. Swets, J. Weng, Using discriminant eigenfeatures for image retrieval, IEEE Trans. Patt. Anal. Mach. Intel. 18 (8) (1996) 831–836.
[48] X. Liu, T. Chen, Pose-robust face recognition using geometry assisted probabilistic modeling, in: Proceedings of IEEE CVPR, 2005, pp. 502509.
[49] Z. Cao, Q. Yin, J. Sun, X. Tang, Face recognition with learning-based descriptor, in: Proceedings of IEEE CVPR, 2010, pp. 2707–2714.
[50] H. Wang, S.Z. Li, Y. Wang, Face Recognition under varying lighting conditions using self quotient image, in: Proceedings of IEEE International Conference of Auto Face Gesture Recognition, 2004, pp. 819–824.
[51] H.F. Chen, P.N. Belhumeur, D.W. Jacobs, In search of illumination invariance, in: Proceedings of IEEE CVPR, 2000, pp. 254–261.
[52] W. Zhang, S. Shan, W. Gao, X. Chen, H. Zhang, Local Gabor binary pattern histogram sequence (LGBPHS): a novel non-statistical model for face representation and recognition, in: Proceedings of IEEE ICCV, 2005, pp. 786–791.