INFORMATICA, 2019, Vol. 30, No. 4, 647–670

DOI: http://dx.doi.org/10.15388/Informatica.2019.223

Multi-Pose Face Recognition Using Pairwise Supervised Dictionary Learning

Ali FARAHANI, Hadis MOHSENI*

Department of Computer Engineering, Shahid Bahonar University of Kerman, Pazhouhesh Square, Kerman, Iran
e-mail: cpt.mazi@gmail.com, hmohseni@uk.ac.ir

Received: March 2018; accepted: April 2019

Abstract. A major challenge in face recognition is handling large pose variations. Here, we proposed to tackle this challenge by a three step sparse representation based method: estimating the pose of an unseen non-frontal face image, generating its virtual frontal view using learned view-dependent dictionaries, and classifying the generated frontal view. It is assumed that for a specific identity, the representation coefficients based on the view dictionary are invariant to pose and view-dependent frontal view generation transformations are learned based on pair-wise supervised dictionary learning. Experiments conducted on FERET and CMU-PIE face databases depict the efficacy of the proposed method.

Key words: face recognition, multi-pose, sparse representation, supervised dictionary learning.

1. Introduction

Face recognition is one of the most important biometric techniques that has clear advantages over other biometric techniques, e.g. it is non-intrusive, natural and passive, where other biometric techniques such as fingerprint recognition and iris recognition require cooperative subjects. To enjoy the non-intrusive nature of face recognition, a system should be able to recognize a face in an uncontrolled environment and an arbitrary situation without the notice of the subject (Zhang and Gao, 2009). This brings serious challenges to face recognition techniques due to this generality in the environment and situation. Conducted evaluations on state-of-the-art face recognition techniques during the past several years, such as the FERET evaluation and the FAT 2004 (Phillips *et al.*, 2000; Messer *et al.*, 2004), have confirmed that there are several major challenges in current face recognition systems which are age, pose, illumination, expression, size, etc. variations. Also, face occlusion with hair, sunglasses, make-up, etc. can bring inconvenience for face recognition techniques. Although most of current face recognition techniques work well under constrained conditions, they fail under uncontrolled cases (e.g. outdoor with uncooperative subjects), since they are sensitive to the mentioned variations on face image.

*Corresponding author.

One of the interesting facts about face images is that face changes occurred by pose variation in images of one identity may be larger than face changes occurred by identity variation in a fixed pose. Face images of different identities in the same pose resemble each other and the differences between them are subtle (Nastar and Mitschke, 1998). This proves the difficulty of multi-pose face recognition, which is a bottleneck for most current face recognition technologies. Therefore, among all variations on the face, in this paper, we concentrate on pose variation challenge. However, as the experimental results will show, the proposed method can tolerate, to some extent, the illumination variation in face images, too.

We organized the rest of the paper as follows. Section 2 reviews some related works in literature for face recognition application. Section 3 reviews related concepts on sparse representation based classification. The proposed method for multi-pose face recognition is presented in Section 4. Extensive experiments are carried out in Section 5 and the experimental results are compared with the results from remarkable algorithms developed before in the literature. Finally, we conclude this paper in Section 6.

2. Related Works

As mentioned in Introduction, one major problem in multi-view face recognition is that the variation in pose may cause changes in face image that are larger than that caused by variation in identity. There are many methods that perform well when training and testing face images are within similar condition and poses, but due to the mentioned difficulties, fewer methods have been proposed in handling the problem of recognizing faces in arbitrary poses.

Multi-Pose face recognition approaches that already have been proposed in the literature can be summarized in two main categories as follows (Zhang *et al.*, 2013):

1. Multi-view approaches that expend the training and/or testing set to encompass more face poses to form a relatively more robust feature set.
2. Invariant approaches which perform some particular transformation to eliminate the variation caused by pose change or to reduce its adverse effect on the final recognition.

Among the methods of the first category, the view-based methods are widely used (Murase and Nayar, 1993; Pentl *et al.*, 1994; Mckenna *et al.*, 1996; Zhou *et al.*, 2001). For instance, view-based Eigenface was proposed to extend the Eigenface to handle the pose problem. One disadvantage of view-based methods is that for each subject, these methods usually require multiple face images with different poses, which is infeasible in real-world applications.

Gross *et al*. proposed the Eigen Light Field (ELF) (Gross *et al.*, 2004) method to tackle the pose problem. ELF first estimates the ELF's of the identity's head from the input images. Then, the test and gallery images are matched by comparing the ELF coefficients. Compared to view-based methods, ELF needs an extra independent training set (different

from the gallery) that contains multiple images of varying poses for each subject. However, in the recognition stage, one face is recognized even if he/she has only one image in the gallery. Providing additional images in different poses brought more depth information of the human face structure, and consequently results in better-reconstructed models compared to the models that use only a single gallery image. However, this method puts a restriction on data collections requirements, because many existing face databases might only contain a few (even single) gallery images (Zhou *et al.*, 2001).

Pentl *et al.* (1994) proposed tied factor analysis model (TFA) to describe pose variation on face images and achieved state-of-the-art face recognition performance under large pose variations. They made an assumption that each identity can be described by an identity vector and images of a single identity in different poses can be generated using this identity vector. This is done by performing identity-independent (but pose-dependent) linear transformation. The identity vectors and the parameters of the linear transformation are estimated using a set of training images in different known poses and the EM algorithm. In fact, TFA searches the transformations to achieve feature extractions that are pose-independent. However, as just linear transformations are considered due to computational feasibility, it could not properly describe pose variations for 2D mapped face images which are non-linear transformations (Zhang and Gao, 2009).

From the second category of multi-view face recognition approaches, one can mention the methods that generate virtual views. In these methods, all face images are normalized to a pre-defined pose (e.g. frontal pose) or the gallery is expanded to cover large pose variations by generating virtual views. As changing pose causes variations that are closely related to the 3D structure of the face, it is a natural idea to build a 3D model from the 2D input face image (Chai *et al.*, 2007). For instance, multi-level quadratic variation minimization (MQVM) (Zhang *et al.*, 2008) uses two gallery images of the frontal view and side view to reconstruct 3D human face for recognition. One of the most successful methods for 3D face model recovery is 3D Morphable Model (3DMM) (Blanz and Vetter, 2003). In this method, PCA is used to model the prior knowledge of face shape and texture. Then any unseen face can be modelled by the linear combination of the prototypes, in which the corresponding shape and texture are expressed by the exemplar faces. The specified 3D face can be recovered from one or more images by optimizing the shape, texture and mapping parameters through an analysis-by-synthesis strategy. However, 3DMM is time-consuming for most real-world applications. To reduce the complexity, Jiang *et al.* (2005) proposed a simplified version of 3DMM to reconstruct the specified 3D face from a single frontal view. They used facial features to reconstruct more efficient personalized 3D face models and their method is based on the automatic detection of facial features on the frontal view.

Unlike 3D model-based approaches, learning-based approaches generally try to learn how to estimate a virtual view directly in 2D space. In Lee and Kim (2006), a method is proposed to generate frontal view face images using a linear transformation in feature space. Features are extracted from non-frontal face images using kernel PCA and then, a transformation from non-frontal view face image to its corresponding frontal view is applied. The transformation is obtained by a least-squared error learning process. As another example of learning-based methods, the Active Appearance Model (AAM) (Cootes

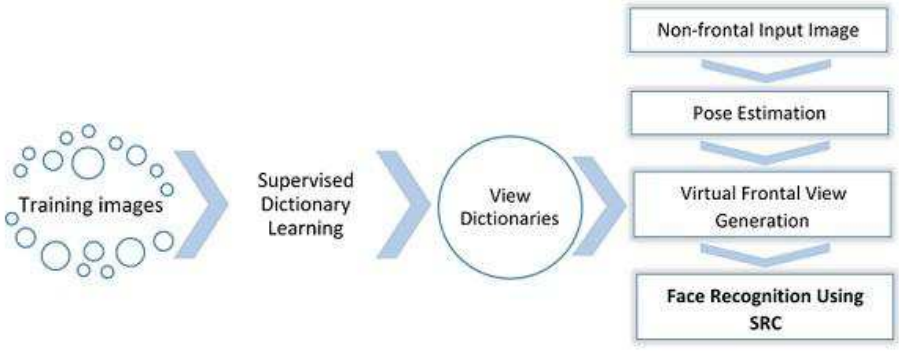

Fig. 1. Flowchart of the proposed method.

et al., 2001) fits an input face image to the pre-learned face model, which consists of separated shape and appearance models. Beymer (1994) proposed a parallel deformation to generate virtual views covering a set of possible poses from a single example view using feature-based 2D wrapping. In this method, a 2D transformation from a standard pose to a target pose is learned. To synthesize a virtual view of gallery faces in the same target pose, the real view in the standard pose is parallel deformed based on the learned 2D transformation on the prototype face.

Another class of methods lying in the second category of multi-view face recognition approaches are subspace-based methods such as Belhumeur *et al.* (1997). These methods seek for the most representative subspace for dimension reduction and feature extraction. Fisherface approach is applied to expressly provide the discrimination among classes when multiple training data per class are available. Through the training process, the ratio of the between-class difference to the within-class difference is to be maximized to find a base of vectors that best discriminate the classes. In Modular PCA (MPCA) (Gottumukkal and Asari, 2004) face images are divided into smaller regions and the PCA approach is applied to each of these regions. Since some of the local facial features of a subject do not vary in pose variation, it is expected that the MPCA be able to cope with pose variation.

The nearest subspace (NS) is a method of the second category that generalizes NN (Nearest Neighbour) method in the sense it classifies the test sample based on the best linear representation of all training samples in each class. The sparse representation based classification (SRC) method (Wright *et al.*, 2009) is a further generalization of NS by representing the test sample using the training samples selected from all samples, both within and across different classes.

In this paper, we propose a novel multi-pose face recognition method by formulating virtual frontal view generation via sparse representation as a prediction problem. The proposed method lies in the second category of multi-pose face recognition. Figure 1 shows a flowchart of the proposed method. As shown in Fig. 1, the proposed method includes 3 steps:

1. Pose estimation step: in many experiments, it has been proved that knowing the pose of the unseen face image can be helpful in recognizing its identity. Contrary

to most face recognition methods that assume prior knowledge about the pose, the proposed method estimates the pose of face image using the sparse representation idea. The proposed pose estimation method is based on the assumption that sparse representation coefficients of different identities in the same pose are closer than representation coefficients of images of the same identity in different poses. Therefore, using the sparse representation of unseen face image over training images of the same pose and repeating this for all poses, one can estimate the pose based on minimizing the reconstruction error on different poses.

2. Virtual frontal view generation step: according to the estimated pose in the previous step, a non-linear mapping is applied to the non-frontal face image in order to generate its virtual frontal view image. This virtual frontal view is then used for the aim of classification. The mapping used in this step is based on sparse representation and the supervised dictionary learning concept. In fact, this step aims to learn view-dependent dictionaries which will be used in the generation of the virtual frontal image from a specific view.
3. Classification (recognition) step: in this step, recognition of the unseen face image is done based on its virtual generated frontal view and an SRC-based classifier.

Therefore, all three steps of the proposed method are based on sparse representation which could be considered as an advantage of the proposed method. Extensive experiments have been conducted on the CMU-PIE (Sim *et al.*, 2002) and FERET (Phillips *et al.*, 2000) face databases to evaluate the efficacy of the proposed method. Section 3 will explain the basic concepts of sparse representation which is the main component in all steps of the proposed method.

3. Sparse Representation

Sparse Coding or Sparse Representation (SR) is a powerful tool in high-dimensional signal processing which has shown strong performance in applications of computer vision, especially face recognition (Wright *et al.*, 2010). It uses a dictionary of base functions (atoms), so each input signal can be approximated by a linear combination of just a sparse subset of atoms. It should be noted that based on the sparse representation theory, similar signals in the same class are expected to be approximated by a similar subset of atoms.

Suppose that there are N training samples from C different classes that are arranged in matrix $A = [a_1, a_2, \ldots, a_N] \in R^{(d \times N)}$ where each sample has d features and the label vector $L = [l_1, l_2, \ldots, l_N]$, $l_i \in [1, \ldots, C]$ stores the label of samples. The sparse representation of test sample $y \in R^d$ over training samples A can be obtained by solving Eq. (1) (Elad, 2010).

$$\hat{x} = \arg\min_x \|y - Ax\|_2^2 + \lambda \|x\|_1, \tag{1}$$

where $\|x\|_1$ is the l_1-norm of x and it is a measure of sparsity and $\|x\|_2$ is the l_2-norm of x. λ is the Lagrangian coefficient and regularizes the pressure of sparsity of x and

corresponding reconstruction error in the first term. Eq. (1) is a relaxed version of a non-convex and an NP-Hard problem that uses l_0-norm instead of l_1-norm.

If sparse representation is used for classification aim (SRC), the class label of test sample y can be obtained based on the minimum reconstruction error criteria as follows:

$$\hat{c} = \arg\min_c \left\| y - AZ_c(x) \right\|_2, \tag{2}$$

where $Z_c(x) : \Re^N \to \Re^N$ is a selection operator that selects coefficients associated with class c from vector x and sets other coefficients to zero.

As already mentioned, a dictionary is a set of basis data (atoms) based on which the sparse representation is obtained. Dictionary atoms can be chosen from raw training data or pre-constructed dictionaries, such as undecimated wavelets, contourlets, curvelets, and more. Although pre-constructed dictionaries result in fast transforms, they are usually limited in sparsifying the signals that they are designed to represent. Alternatively, one can use learning methods to obtain a tunable dictionary in which each atom is generated by controlling some parameters. This is called Supervised Dictionary Learning (SDL) and learned dictionaries are expected to be able to adapt to different input samples (Elad, 2010). The proposed method uses both pre-constructed dictionaries and supervised dictionary learning to have a proper dictionary in each step and the best performance in general.

The next section will explain the proposed method and how the sparse representation concept can be extended to have good performance on multi-pose face recognition.

4. Proposed Method

In this section, a sparse representation based multi-pose face recognition method is proposed which consists of three main steps:

1. Estimating the pose of a given face image based on SRC.
2. Generating a virtual frontal view of the given face image according to the estimated pose and the learned SR-based non-frontal to frontal view mapping.
3. Recognizing the face image using the generated frontal view and SRC.

The most important step in the proposed method is generating a virtual frontal view of a non-frontal face image. Since the proposed method learns view-dependent transformations to map the non-frontal face image to the frontal one, choosing the proper transformation needs the pose of the face image. Therefore, the first step of the proposed method is devoted to pose estimation. According to the estimated pose, a non-frontal to frontal view mapping is applied which generates a virtual frontal face image. Having the virtual frontal view in hand, SRC is used for the aim of face recognition. The following subsections explain these three steps of the proposed method in more details.

4.1. *Pose Estimation Based on SRC*

As mentioned earlier, prior knowledge of the pose of a face is an essential information in many face recognition techniques. It is often beneficial if the pose angle of the input face image can be estimated before recognition such as in modular PCA (MPCA) (Pentl *et al.*, 1994) and eigen light-field (Gross *et al.*, 2004). There are many efforts for automatic pose estimation in the literature. As the focus of this paper is on face recognition and not pose estimation, an interested reader is referred to Murphy-Chutorian and Trivedi (2009) and Ding and Tai (2016) as good surveys on face pose estimation.

It is obvious that two face images from different identities in the same pose are visually more similar than two face images of an identity in different poses. This can be used as a clue for face pose estimation aim, a face image in a specific pose can be estimated by a linear combination of face images of other subjects in the same pose. The proposed pose estimation method is based on the assumption that sparse representation coefficients of different identities in the same pose are closer than representation coefficients of images of the same identity in different poses. Therefore, using the sparse representation of unseen face image over training images of the same pose and repeating this for all poses, one can estimate the pose based on minimizing the reconstruction error on different poses. A similar idea has been used in Yu and Liu (2014) where it is assumed that a face image in a specific pose cannot be approximated by a combination of face images in other poses.

Suppose there are P classes of different poses A_p, $p = 1, \ldots, P$. The p-th class $A_p = [a_p^1, a_p^2, \ldots, a_p^{n_p}] \in \Re^{(dn_p)}$ is called the view dictionary of pose p that has n_p training face images from different identities in this pose and d is the dimension of each face image. Based on SR theory, every unseen face image $y \in \Re^d$ is expected to be expressed as a sparse representation of images in matrix A_p in a particular pose. The sparse coefficients of image y over the view dictionary A_p is the $\hat{x}_p$ vector that can be obtained as follows:

$$\hat{x}_p = \arg\min_{x_p} \|y - A_p x_p\|_2^2 + \lambda \|x_p\|_1, \tag{3}$$

where λ is the regularization parameter as before. Therefore, face image y is reconstructed based on different view dictionaries. The view dictionary that reconstructs the face image with minimum error determines the pose of the face image y. In other words, the pose of face image y is estimated based on minimizing the reconstruction error among all view dictionaries:

$$\hat{p} = \min_p \|y - A_p x_p\|_2, \quad p = 1, \ldots, P, \tag{4}$$

where $\hat{p}$ is the estimated pose. Actually, this shows that $A_{\hat{p}}$ is the best view dictionary that can reconstruct the input face image from a linear combination of its set of face images. The proposed pose estimation algorithm is summarized in Algorithm 1. Figure 2 represents an example of the proposed pose estimation method where there are seven different poses (seven view dictionaries).

Figure 2(a) shows the input face image on the top and the 7-th reconstructed face images with respect to seven view dictionaries below that. As it is obvious, reconstructed

Algorithm 1: Sparse representation based pose estimation.

Input: Training View Dictionaries $\{A_p\}$ $p = 1, \ldots, P$, non-frontal input face image y.

Output: Estimated pose $\hat{p}$

Steps:

1. Calculate the sparse representations of y over A_p for $p = 1, 2, \ldots, P$ Eq. (3) to obtain sparse representation vectors $\hat{x}_1, \hat{x}_2, \ldots, \hat{x}_p$.

2. Estimate the pose of input face image via Eq. (4).

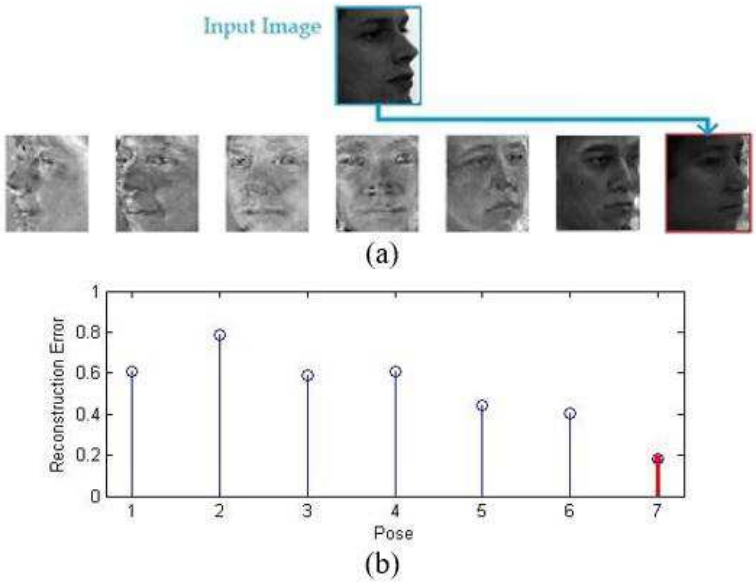

Fig. 2. An example of pose estimation based on sparse representation. (a) reconstructed input face image over 7 different view dictionaries A_p, $p = 1, \ldots, 7$ and (b) reconstruction error for each pose. Reconstruction error for 7-th pose is minimum and the input face image is supposed to be in this pose.

image from the last dictionary (A_7) seems to be the most similar one to the input face image, where the reconstruction error plot in Fig. 2(b) confirms this consequence. Thus, the input face image is supposed to be in the last pose.

The proposed pose estimation method has some advantages over many other pose estimation methods. First, there is no assumption on the number of training face images in each pose and view dictionaries can have a different number of atoms. Also, no feature selection or 3D model of the face is required for pose estimation, so no image registration and heavy computation is used. However, the main shortcoming of the proposed pose estimation method is its accuracy drop in small pose intervals which will be discussed more in Section 5.2.

4.2. *Virtual Frontal View Generation*

In many face recognition methods, one of the key steps for achieving multi-pose face recognition is pose normalization or virtual frontal view generation. Obviously, a frontal face image contains the most details of the face which are beneficial for face recognition, compared to a non-frontal face image. In order to compensate for the loss of details in non-frontal views, one can try to generate a virtual frontal view from a non-frontal view. In this paper, this task is formulated as a general prediction framework which predicts a mapping from each non-frontal view to the frontal view, where the mapping is identity-independent. The purpose of this mapping is to estimate a frontal face image $\hat{b}_1 \in \Re^d$ given its non-frontal face image $b_p \in \Re^d$ in pose p. Modelling the virtual frontal view generation with linear mapping is as follows:

$$b_1 = V_p(b_p) = W_p b_p, \tag{5}$$

where $V_p(.)$ is the linear mapping function and W_p is the linear mapping matrix for pose p. Linear mapping function $V_p(.)$ can be achieved via a learning process. GLR and LLR (Chai *et al.*, 2007) are general least square problems that use regression-based methods to find a good mapping. Another idea to find the mapping function is introduced in LSRR (Zhang *et al.*, 2013) which assumes that the face images of one identity observed from different views share the same sparse representation coefficients over different view dictionaries. In other words, suppose f_1 and f_2 are two face images of one identity in poses p_1 and p_2 , respectively. The sparse representation coefficients of these two face images over view-dependent dictionaries related to p_1 and p_2 poses are supposed to be similar. Therefore, if sparse representation coefficients of a non-frontal face image of an identity are available over its view dictionary, these coefficients can be used to generate the virtual frontal view of that identity, using the frontal view dictionary. Consequently, considering face images of the i-th identity, we have the following set of equations:

$$\begin{cases} b_1^i = A_1 x^i + e_1 \\ \quad \vdots \\ b_p^i = A_p x^i + e_p \\ \quad \vdots \\ b_P^i = A_P x^i + e_P, \end{cases} \tag{6}$$

where A_p is the view dictionary of pose p, b_p^i is the face image of identity i in pose p and e_p is the reconstruction error in pose p. The sparse representation coefficients x^i are shared among all the P views of the i-th identity. These equations say that the face image from pose p can be generated from sparse representation coefficients x^i with the corresponding view dictionary A_p. Therefore, the key of virtual view generation lies in the recovery of the sparse representation coefficients x_i. The idea of sharing the sparse representation coefficients among different poses somehow reminds the idea used in Prince *et al.* (2008) where the authors assumed a face manifold and an identity space (latent space)

and declared that the representation of each identity does not vary with pose. As another example, one can mention the research done in Sharma *et al.* (2012) which aims to find the sets of projection directions for different poses such that the projected images of the same identity in different poses are maximally correlated in the latent space.

Based on the discussion above, given training samples arranged in different view dictionaries A_p $(p = 1, \ldots, P)$, the non-frontal to frontal mapping function for input face images in pose p (b_ps) can be obtained by first finding the sparse representation coefficients of b_p on view dictionary of pose p, then utilizing these coefficients with the frontal view dictionary A_1 as follows:

$$\hat{x}_p = \arg\min_{x_p} \|b_p - A_p x_p\|_2^2 + \lambda \|x_p\|_1, \tag{7}$$

$$\hat{b}_1 = V_p(b_p) = A_1 \hat{x}_p, \tag{8}$$

where $\hat{x}_p$ is the best vector of sparse representation coefficients of b_p over view dictionary A_p, λ is the regularization parameter as mentioned before and $\hat{b}_1$ is the virtual frontal view corresponding to the non-frontal view b_p.

It is worth noticing that the mapping in Eq. (8) is based on two parameters, view dictionaries and sparse representation coefficients. As the sparse representation coefficients are obtained via an optimization problem based on a view dictionary, one of the factors that play an important role in high accuracy mapping is the selection of view dictionaries. These dictionaries can be simply made form training images of each pose or can be learned more effectively in a dictionary learning process. As training images are accompanied by identity label, using them as view dictionary atoms might not successfully generate a face image from a new identity. In other words, identity-independent view dictionaries are expected to be more efficient in generating face images from new identities. Therefore, one of the key steps for increasing the accuracy of the proposed method in obtaining the sparse representation coefficients and generating the virtual frontal view is to learn desirable identity-independent view dictionaries. The next subsection explains a supervised dictionary learning process to learn A_ps as efficient as possible.

4.3. *Supervised View Dictionary Learning*

Suppose that b_p and b_1 are two face images of one identity in non-frontal pose p and frontal pose 1, respectively, $\hat{x}_p$ is the sparse representation of b_p over view dictionary A_p, and $\hat{x}_1$ is the sparse representation of b_1 over view dictionary A_1. As mentioned in the previous section, view dictionaries A_p and A_1 are called desirable if the sparse representations $\hat{x}_p$ and $\hat{x}_1$ are the same or at least close enough. So, the aim of this subsection is to learn view dictionaries that share similar sparse coefficients for face images of one identity in different poses. This is achieved via a supervised view dictionary learning process.

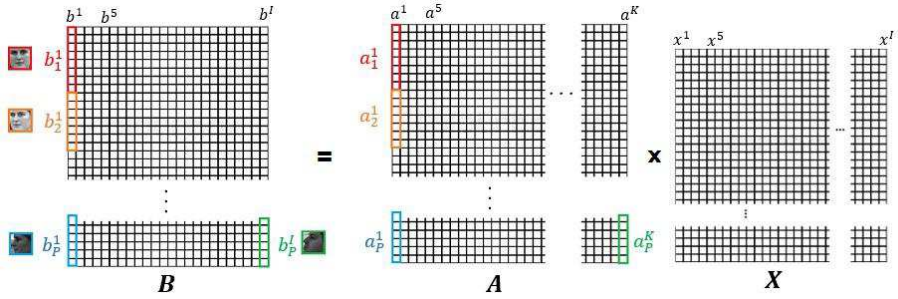

Fig. 3. Example of sparse dictionary learning. $B \in \Re^{(dP\times I)}$ is the learning face images in P different poses for I identities, $A \in \Re^{(dP\times K)}$ is the dictionary includes P different view dictionaries which each view dictionary can be extracted by separating d rows related to each pose, and $X \in \Re^{(K\times I)}$ is the sparse representation matrix.

By concatenating the P equations in Eq. (6), while omitting the identity parameter i for simplicity, we have:

$$\begin{pmatrix} b_1 \\ \vdots \\ b_p \\ \vdots \\ b_P \end{pmatrix} = \begin{pmatrix} A_1 \\ \vdots \\ A_p \\ \vdots \\ A_P \end{pmatrix} x + \begin{pmatrix} e_1 \\ \vdots \\ e_p \\ \vdots \\ e_P \end{pmatrix} \rightarrow B = Ax + e \qquad (9)$$

subject to $\|x\|_1 \leqslant C$,

which states that the P views, when concatenated together, should have the same sparse representation with respect to the concatenated view dictionary. Given the training dataset $\{b_p^i\}_{p=1,\ldots,P}^{i=1,\ldots,I}$, where i is the index for identities and p is the index for poses, the training set is rearranged by concatenating the P views of each identity in $\{b^i\}_{i=1,\ldots,I}$ vector, where $b^i = [b_1^i b_2^i b_P^i]^T \in \Re^{(dp\times 1)}$ and $B \in \Re^{(dP\times I)}$ is the matrix made from concatenating face images of I different identities. Now, the view dictionaries can be learned via the following minimization problem:

$$\langle \hat{A}, \hat{X} \rangle = \arg\min_{A,X} \sum_{i=1}^{I} \left\| b^i - Ax^i \right\|_2^2 + \lambda \left\| x^i \right\|_1, \qquad (10)$$

where $\hat{A} = [\hat{A}_1^T, \hat{A}_2^T, \ldots, \hat{A}_P^T]^T \in \Re^{(dP\times K)}$ is the learned dictionary and $\hat{X} = [\hat{x}^1, \hat{x}^2, \ldots, \hat{x}^I] \in \Re^{(K\times I)}$ is the sparse coefficients matrix whose column i is the sparse representation vectors of the training samples in b^i and K is the dictionary size. Eq. (10) aims to jointly find the proper sparse representation coefficients and the dictionary. It describes face images from the ith identity (b^i) as the sparsest representation x^i over dictionary A. After $\hat{A}$ is learned, the view dictionaries $\{\hat{A}^p\}_{p=1,\ldots,P}$ are obtained by splitting $\hat{A}$ into P parts, e.g. view dictionary of pose p ($\hat{A}_p$) can be achieved by separating d rows of $\hat{A}$ that are corresponding to the p-th pose (rows $pd - d + 1$ to dp). Figure 3 demonstrates these matrices and the dictionary learning process visually.

In order to properly choose the dictionary size K, it is worthy to remind some points on dictionary characteristics. Dictionary $A \in \Re^{(dP \times K)}$ is considered undercomplete if $K < dP$ or overcomplete if $K > dP$. When $(K = dP)$, dictionary is considered as a complete dictionary. From a representational point of view, a complete dictionary does not help in any improvement and is neglected. Undercomplete dictionaries are strongly related to dimensionality reduction. Principal component analysis is a famous example of this case where dictionary atoms have to be orthogonal. However, putting orthogonality constraint on dictionary atoms limits the choice of atoms which is the main disadvantage of undercomplete dictionaries. On the other side, overcomplete dictionaries do not have the orthogonality constraint, therefore, they allow for more flexible dictionaries and richer data representation (Elad, 2010).

Although all view dictionaries can be learned simultaneously using Eq. (10), the learning process will be impractical for large dictionary sizes or high dimensional data. Consider a situation where each identity has images in 10 poses and each image has about 1000 pixels (a small 30×35 face image), so each column of dictionary has about 10000 entries. If the dictionary size is adjusted to 1000 atoms, the size of dictionary will be 10000×1000. Doing computation on a dictionary of this huge size is impractical because of memory and computational limitations. To dominate this problem, in this paper, pairwise dictionary learning is proposed, where each view dictionary is learned separately. In other words, in order to learn the view dictionary for pose p, the training matrix will be $B_p \in \Re^{(2d \times I)}$ where each column of B_p is $[b_p^i b_1^i]^T \in \Re^{(2d \times 1)}$ (face images of frontal pose and images of non-frontal pose p). In this case, the optimization in Eq. (10) results in $\hat{A} \in \Re^{(2d \times K)}$ where the first d rows of $\hat{A}$ can be considered as learned view dictionary for pose p. It should be noted that view dictionaries are not learned simultaneously in pairwise dictionary learning. However, as training images in frontal pose are the same for learning all view dictionaries, it is expected that describing face images of one identity in different poses by these view dictionaries has similar sparse representation coefficients. Figure 4 shows the effect of dictionary learning on sparse representation of three face images of an identity in different poses. The first column shows the sparse coefficients when dictionary atoms are simply the training data in different poses while the second column shows sparse coefficients obtained based on learned dictionary. As expected, the representation coefficients of the three images shown in the second column are more similar compared to the ones in the first column. This observation confirms the effect of dictionary learning on unifying the sparse representation coefficients of face images of one identity in different poses. Also, the figure depicts the increase in sparsity of coefficients after dictionary learning, which is another aim of dictionary learning. Getting back to dictionary learning process in Eq. (10), several dictionary learning methods have been proposed since now that can be divided into two groups: 1) unsupervised dictionary learning methods such as MOD (Engan *et al.*, 1999) and K-SVD (Aharon *et al.*, 2006) and 2) supervised dictionary learning methods such as SDL (Mairal *et al.*, 2009) and LCKSVD (Jiang *et al.*, 2013). The K-SVD method is introduced to efficiently learn an overcomplete dictionary and has been successfully applied to image restoration and image compression. K-SVD focuses on the representational power of the learned dictionary, but does not consider the discrimination

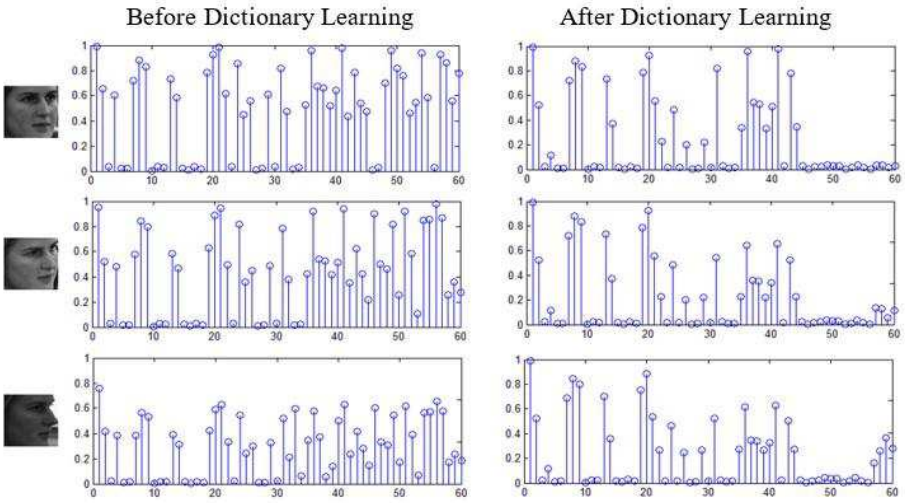

Fig. 4. Effect of dictionary learning on sparse representations of 3 different pose images of one identity. The first and second columns show the sparse representation of face images using training samples and the learned dictionary, respectively. As expected, representation coefficients in second column are more similar in different poses, while coefficients are sparser in each pose.

capability of it. LCKSVD (Jiang *et al.*, 2013) is a supervised extension of K-SVD that uses supervised information (labels) of training samples to learn a compact and discriminative dictionary. As LCKSVD has proved itself as a successful supervised dictionary learning method, it has been used here for learning view dictionaries. The objective function defined by LCKSVD is as follows:

$$\langle \hat{A}, \hat{X}, \hat{T} \rangle = \arg \min_{A,X,T} \|B - AX\|_F^2 + \alpha \|Q - TX\|_F^2 \quad \text{s.t. } \forall i,\ \|x^i\|_0 \leqslant C, \tag{11}$$

where $\|.\|_F$ is the Frobenius norm and B, A and X are training data, dictionary and sparse coefficients matrices, respectively. $Q = [q^1, q^2, \ldots, q^I] \in \Re^{(K \times I)}$ is the discriminative sparse code of training samples in B and is initialized based on the labels of training samples and desired labels of dictionary atoms. For example, if i-th atom of dictionary has the same label as j-th training sample, then $Q(i, j) = 1$, else $Q(i, j) = 0$. T is a linear transformation matrix and the term $\|Q - TX\|_F^2$ enforces the sparse coefficients X to approximate the discriminative sparse codes Q. This term enforces the samples from the same class to have very similar sparse representations. The first and second terms of Eq. (11) are the reconstruction error and the discrimination power, respectively, where α controls the contribution between these two terms. The implementation of LCKSVD is available by the LCKSVD authors and is used in this paper for solving Eq. (11). For more details on LCKSVD, we refer the interested reader to Jiang *et al.* (2013).

After obtaining view dictionaries, virtual frontal view generation can be done by first estimating the pose $\hat{p}$ of the input face image y by Algorithm 1, then finding $\hat{x}_{\hat{p}}$ as the sparse representation of y over the learned view dictionary of the estimated pose. Finally,

Algorithm 2: Virtual View Generation

Input: Training samples $\{b_p^i\}_{p=1,\ldots,P}^{i=1,\ldots,I}$ for view dictionary learning, non-frontal input face image y.
Output: Virtual frontal view $\hat{y}_1$
Steps:
1. Perform view dictionary learning using LCKSVD via Eq. (11) to obtain $\hat{A}_p$, $p = 1, \ldots, P$.
2. Estimate the pose of y using using Algorithm 1 and obtain $\hat{p}$.
3. Solve sparse representation of y over $\hat{A}_{\hat{p}}$ and obtain $\hat{x}_{\hat{p}}$ via Eq. (3).
4. Generate Virtual frontal view $\hat{y}_1 = \hat{A}_1 \hat{x}_{\hat{p}}$.

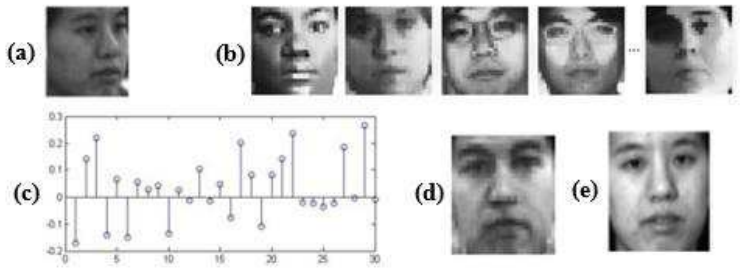

Fig. 5. Virtual frontal view generation. (a) non-frontal input face image y_p, (b) learned view dictionary for pose p ($\hat{A}_p$), (c) sparse representation of y_p over $\hat{A}_p$, (d) generated virtual frontal view $\hat{y}_1$, (e) actual frontal view of input face image y_1.

the virtual frontal view of y is generated by multiplying the sparse representation $\hat{x}_{\hat{p}}$ to the learned view dictionary of frontal pose $\hat{A}_1$. The algorithm of virtual view generation is summarized in Algorithm 2 and an example of virtual view generation is shown in Fig. 5.

4.4. *Multi-Pose Face Recognition*

The previous subsection explained the proposed virtual frontal view generation algorithm which was based on supervised dictionary learning. This subsection completes the previous steps by recognizing the generated frontal view. As mentioned earlier, SRC method has shown superior performance on frontal view face recognition, therefore, in this paper, SRC is used as the classifier in the recognition step. The overall view of the 3 steps of the proposed method is shown in Algorithm 3.

5. Experimental Results

The proposed method is evaluated on CMU-PIE (Sim *et al.*, 2002) and FERET (Phillips *et al.*, 2000) face databases and three experiments are carried out to show the effective-

Algorithm 3: Multi-Pose Face Recognition with Supervised Dictionary Learning (MPSDL)

Input: Frontal and non-frontal Training face images $\{b_p^i\}_{p=1,\ldots,P}^{i=1,\ldots,I}$, non-frontal input test image y.
Output: class label of y
Steps:
1. Generate virtual frontal view of y using Algorithm 2 and obtain $\hat{y}_1$.
2. Solve sparse representation of $\hat{y}_1$ over frontal training face images via Eq. (1).
3. Class label of y is obtained by minimum reconstruction error criterion via Eq. (2).

Fig. 6. Different poses of a subject in CMU-PIE face database (Sim *et al.*, 2002).

ness of the proposed method. Section 5.1 explains the databases and how to prepare the face images for experiments. In Section 5.2, performance of the proposed pose estimation method is measured in both small and large pose variations. Section 5.3 considers the virtual view generation step and compares the generated frontal faces with similar view generation methods such as GLR and SRR. Finally, in Section 5.4, the accuracy of the proposed multi-pose face recognition is evaluated based on virtual frontal views.

5.1. *Databases for Evaluations*

CMU-PIE and FERET databases contain a large number of face images in different illumination, viewpoints and expressions. CMU-PIE database has 68 identities who were imaged under 13 different poses, 43 different illumination conditions and 4 different expressions. Figure 6 shows the variation of poses in the CMU-PIE database, images are within $-90°$ to $+90°$ with $\pm 22.5°$ interval in yaw and about $\pm 20°$ in pitch. FERET database contains more than 1000 identities in different conditions. From them, 200 subjects have all 9 different pose variations within $\pm 60°$ in yaw ($0°$ in pitch). Specifically, the poses are $\pm 60°$, $\pm 40°$, $\pm 25°$, $\pm 15°$ and frontal pose $0°$. Figure 7 shows the variation of poses in the FERET database. In our experiments, 5 poses of CMU-PIE database are used in $90°$, $67.5°$, $45°$, $22.5°$, $0°$ with 16 different illumination conditions and neutral expression. From FERET database, face images of poses $60°$, $40°$, $15°$ and $0°$ are selected in neutral

Fig. 7. Different images from FERET database (Phillips *et al.*, 2000).

Table 1
Average accuracy of pose estimation for 5 poses (22.5° pose interval) on CMU-PIE.

Pose	0°	22.5°	45°	67.5°	90°	Average
Accuracy (%)	96.5	90.6	89.3	91.1	92.7	92.0

Table 2
Average accuracy of pose estimation for 3 poses with larger pose interval (45° pose interval) on CMU-PIE.

Pose	0°	45°	90°	Average
Accuracy (%)	99.12	98.25	98.77	98.6

expression and illumination condition. For pre-processing, all images are cropped manually (such that eyes and mouth level are fixed), resized to 28×28 pixels and histogram equalization is performed on them.

5.2. *Pose Estimation*

As mentioned in Section 4, it is necessary to know the pose of a non-frontal face image in order to generate its virtual frontal view, because it is necessary to select the proper view dictionary. Table 1 and Table 2 show the accuracy of pose estimation algorithm for 5 poses (22.5° pose interval) and 3 poses (45° pose interval) of CMUPIE database. In each pose, 10 face images from 68 identities are randomly selected to construct the view dictionary. The remaining 58 images per pose are used for evaluation. The accuracies reported in these two tables are obtained by averaging several runs.

Both Tables 1 and 2 represent high accuracy (above 90%) in pose estimation, which is acceptable for many applications. The comparison of the results from these two tables implies that the accuracy of pose estimation increases when the difference between adjacent poses (pose interval) increases. This is because the dictionary atoms of different poses are more distinct and adjacent poses are more discriminable in large pose intervals. Although there might be methods in pose estimation with higher accuracies in small pose intervals, the proposed pose estimation method is simple, fast and accurate enough for the aim of virtual view generation and face recognition. Also, it is based on sparse representation which is the base for the other two steps of the proposed method, too.

5.3. *Virtual Frontal View Generation*

The following experiment assesses the frontal view generation step of the proposed algorithm. Figure 8 shows the virtual frontal views generated from different methods for 22.5°,

Fig. 8. Virtual view generation of different methods for different poses.

45°, 90° poses. As the figure shows, compared to the generated frontal faces by LSRR or GLR methods, generated faces by the proposed method (MPSDL) are visually more similar to the ground-truth images. In fact, generated faces by GLR do not contain much details and generated faces for different identities are similar with artifacts for pose 90°. SRR generated faces have less artifacts and are visually acceptable, but are over-smoothed and details are lost. The LSRR method is similar to SRR, but it performs locally on small patches of an image, so its generated faces have less artifact but are locally smoothed. It requires many overlapped patches to generate detailed images. In contrast, the proposed MPSDL method can generate visually acceptable frontal faces which are similar to ground-truth images and have enough details to be used in recognizing identities.

For evaluating different view generation methods, a 10-fold cross validation strategy is used on CMU-PIE database. In each fold, 61 identities in 16 illuminations are selected for dictionary learning and the remaining 7 identities in 16 different illuminations are used for evaluation. Table 3 shows the Mean Square Error (MSE) between generated frontal views and ground-truths in three poses. The reported results in this table are based on 16 different illuminations of each pose. Chosen values for different parameters are mentioned in caption of the table where $\sigma(B)$ is the standard deviation of matrix B. As can be seen from Table 3, for large pose variations (45° and 90°), the MSE between generated virtual frontal views and the ground-truths is smaller in the proposed method compared to GLR and SRR methods. The reason is that, compared to GLR, the proposed method has no assumption on linear transformation between different poses which is not correct for large pose angles. Also, compared to SRR, the proposed method is based on supervised dictionary learning which uses the discrimination in data and is expected to generate faces with

Table 3
MSE of virtual frontal view generation of different methods ($K = 200$, $\alpha = 1$, $\lambda = \sigma(B)$).

Pose	22.5°	45°	90°	Average
GLR (Chai *et al.*, 2007)	0.3228	0.4725	0.4943	0.4298
SRR (Zhang *et al.*, 2013)	**0.3125**	0.3890	0.4427	0.3814
MPSDL (proposed)	0.3243	**0.3858**	**0.4004**	**0.3701**

more details. Since the proposed method is a holistic approach, LSRR and LLR methods are not included in this comparison, because of their locality manner. Obtained results in Table 3 clearly show that the proposed method can generate accurate virtual views even in large pose variations which can be considered as a desired property of the proposed method. As expected, MSE of different methods is more similar in small pose angles where transformation between poses is nearly linear.

It should be noted that the proposed method aims to generate discriminative virtual frontal views and it does not concentrate on visually good generated faces. However, the results in Table 3 depict that based on MSE, the proposed method can generate more accurate virtual frontal views in large pose angles, compared to other methods.

5.4. *Multi-Pose Face Recognition*

In this section, we evaluate the performance of the proposed multi-pose face recognition method. Virtual view generation can be considered as a preprocessing that is independent of the classification step. Therefore, face recognition based on generated frontal views can be done with various kind of classifiers. Since the SRC method has been successful in frontal face recognition task (Wright *et al.*, 2009), in this paper, SRC is used as a classifier for the generated virtual views. The performance of the proposed method is compared to GLR, SRR and LSRR which are all based on virtual view generation. 10-fold cross validation is used to evaluate the performance of each method on CMU-PIE and FERET databases. For CMU-PIE database, each fold contains 7 identities in 5 poses and 16 different illumination conditions. For FERET database, each fold contains 20 identities in 4 poses and neutral illumination and expression conditions. Using this setting, recognition accuracy of different methods on CMU-PIE and FERET databases are shown in Table 4 and Table 5, respectively. In these tables, it is assumed that the pose angle of test images is known and the proposed pose estimation in Algorithm 1 is not used. In both tables, dictionary size parameter is adjusted to 200 for SRR, LSRR and MPSDL methods. Also, for LSRR, images have been partitioned into 16 patches.

Discussing the results reported in Tables 4 and 5, the recognition accuracy using raw images without virtual view generation will decrease rapidly when the pose angle increases. The GLR method performs better compared to raw images and improves the accuracy, however, its performance is not satisfying for large pose angles because of using linear assumption in virtual view generation. The SRR generally performs better than GLR for large pose angles. Using local patches, LSRR improves the accuracy of SRR, but it seems that both methods suffer from unsupervised dictionary learning. The proposed MPSDL method outperforms other methods and is more robust in large pose variations.

Table 4
Recognition accuracy (%) of different methods (with known pose angle) on CMU-PIE database for 4 poses ($K = 200$, $\alpha = 1$, $\lambda = \sigma(B)$).

Pose	22.5°	45°	67.5°	90°	Average
Raw images	70.3	38.7	21.0	15.5	36.3
GLR (Chai *et al.*, 2007)	87.9	42.9	35.7	32.8	49.8
SRR (Zhang *et al.*, 2013)	90.1	65.6	43.4	40.7	59.9
LSRR (Zhang *et al.*, 2013)	**91.2**	74.9	49.5	48.9	66.1
MPSDL (Proposed)	90.9	**76.2**	**55.8**	**52.1**	**68.7**

Table 5
Recognition accuracy (%) of different methods (with known pose angle) on FERET database for 3 poses ($K = 200$, $\alpha = 1$, $\lambda = \sigma(B)$).

Pose	15°	45°	60°	Average
Raw Images	89.5	27.1	35.2	50.6
GLR (Chai *et al.*, 2007)	86.7	33.7	31.5	50.7
SRR (Zhang *et al.*, 2013)	91.9	52.6	43.0	62.5
LSRR (Zhang *et al.*, 2013)	**94.6**	57.5	45.1	65.7
MPSDL (proposed)	93.7	**62.2**	**50.7**	**68.9**

Table 6
Recognition accuracy (%) of different methods with pose estimation phase on FERET database for 3 poses ($K = 200$, $\alpha = 1$, $\lambda = \sigma(B)$).

Pose	15°	45°	60°	Average
Raw Images	88.9	24.8	35.1	49.6
GLR (Chai *et al.*, 2007)	89.0	31.6	30.9	50.5
SRR (Zhang *et al.*, 2013)	91.6	52.5	43.6	62.5
LSRR (Zhang *et al.*, 2013)	**94.1**	57.2	42.4	64.5
MPSDL (proposed)	93.4	**60.4**	**49.2**	**67.7**

In order to investigate the impact of automatic pose estimation in step 1 on recognition accuracy, Table 6 shows the results of the proposed multi-pose face recognition on FERET database while using Algorithm 1 for pose estimation. As expected, the results in this table are very close to those of Table 5 which confirms the perfect performance of Algorithm 1 for pose estimation. Therefore, the effect of pose estimation error on recognition accuracy can be ignored.

The effect of dictionary size in recognition accuracy has been investigated through experiments done with different dictionary sizes. It should be noted that for raw images and the GLR method that are not based on dictionary learning, dictionary size points to the number of samples in a training set. Table 7 and Fig. 9 show the recognition accuracy of different methods under different dictionary sizes for 22.5° pose of CMU-PIE database. For small dictionary sizes, LSRR method outperforms others because of using small patches.

Increasing the dictionary size, face recognition accuracy increases in all methods of Table 7 (except for GLR), but as can be seen, performance of MPSDL increases more

Table 7
Recognition accuracy (%) for different dictionary sizes ($\alpha = 1$, $\lambda = \sigma(B)$) for 22.5° pose in CMU-PIE.

Dictoinary size	50	100	200	300	500
Raw images	70.2	72.3	70.0	72.5	72.2
GLR (Chai *et al.*, 2007)	59.1	68.9	86.9	53.7	44.1
SRR (Zhang *et al.*, 2013)	76.2	77.4	90.1	79.6	88.7
LSRR (Zhang *et al.*, 2013)	**80.0**	**80.9**	**91.2**	82.6	93.1
MPSDL (proposed)	75.8	78.7	90.9	**92.1**	**94.5**

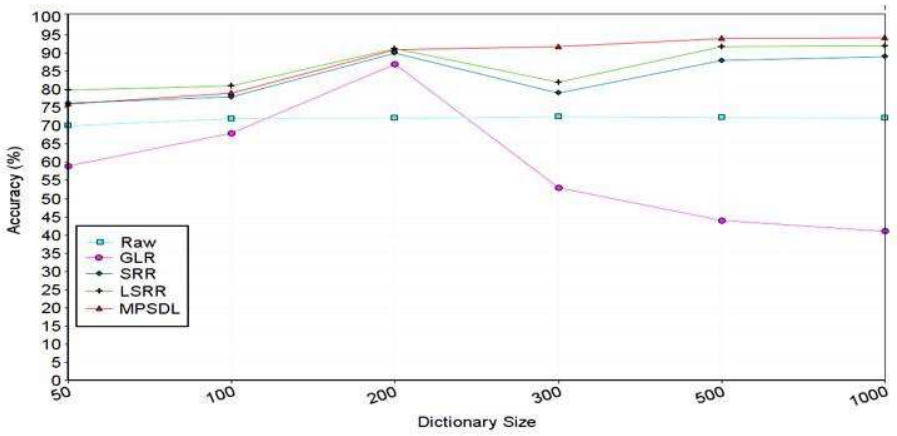

Fig. 9. Comparison of the face recognition accuracy with different dictionary sizes. MPSDL overtakes other methods in larger dictionary sizes.

rapidly. For large dictionary sizes, accuracy of MPSDL overtakes other methods which is because of the use of supervised dictionary learning. Compared to other methods, GLR method shows weak performance for large dictionary sizes, which is because of overfitting of regression based methods in large number of training samples. However, it should be noted that the use of huge databases and big dictionaries might have some computational problems due to the use of huge memory size and very long processing time. One might cope with these problems using high-performance hardware devices and a quantization or clustering method prior to dictionary learning.

Although the proposed method has better performance compared to similar methods, it cannot overtake the state-of-the-art methods which are not based on sparse representation and 2D virtual view generation. For instance, Table 8 shows the comparison between the proposed method and some well-known and state-of-the-art methods in multi-pose face recognition on CMU-PIE database. As can be inferred from this table, 3D reconstruction based methods such as 3DMM (Blanz and Vetter, 2003) and Probabilistic Geometry Assisted FR (Liu and Chen, 2005) can recognize face images with higher accuracy, because they utilize the 3D model of face by aligning 2D images with either a general or an identity specific 3D model which is computationally expensive. TFA (Prince *et al.*, 2008) is a 2D method which benefits from the latent variable view point for face recognition, but

Table 8
Comparison of proposed method and some popular methods on CMU-PIE.

Method	Training images	Accuracy (%)
3DMM (Blanz and Vetter, 2003)	0°, 16°, 60°. 22 illuminations	92.1
PGA FR (Liu and Chen, 2005)	0°, 15°, 30°, 45°, 60°	86.0
TFA (Prince *et al.*, 2008)	0°, 16°, 60°	91.0
LBP (Ahonen *et al.*, 2006)	0°, 30°, 60°	74.2
MPSDL (Proposed)	0°, 22°, 45°, 60°. 16 illuminations.	81.7

it requires two face images (one frontal and one non-frontal) in its gallery for recognition while the proposed method only requires one frontal face image. Therefore, it can be concluded that the recognition accuracy of the proposed method for multi-pose face recognition is persuasive from a computational point of view and when there is only one frontal image of each person in hand. However, based on the idea and obtained results in Zhang *et al.* (2015), Zhao *et al.* (2016), one can extend the proposed method by using mixed norm $l_{(p,q)}$ instead of l_1 norm and hopefully decrease the frontal face generation error while increasing the face recogntion accuracy.

6. Conclusions

In this paper, we proposed a multi pose face recognition method based on sparse representation and supervised dictionary learning. The proposed method generates virtual frontal views from non-frontal views based on the assumption that the images of an identity observed from different views share the same sparse representation coefficients over all view dictionaries. Also, to increase virtual view generation performance and reconstruction accuracy, a supervised dictionary learning is used to generate adapted dictionary atoms. As dictionary learning is usually expensive from computational point of view, the proposed method benefits from pair-wise dictionary learning which learns each view dictionary separately. As a preprocessing step, the proposed method uses a sparse representation based pose estimation, while sparse representation based classification is used for face recognition in the last step. Therefore, all steps of the proposed method are based on sparse representation. Experiments carried out on FERET and CMU-PIE databases show the superior performance of the proposed method compared to other similar methods especially in confronting large pose angles. Compared to the state-of-the art methods, the proposed method has acceptable recognition accuracy from computational point of view while it requires only one frontal image of each subject for recognition. For further work, we would suggest to extend the proposed method to work locally on small patches of a face image, and to investigate how using mixed norms in the objective function can increase the recognition accuracy.

References

Aharon, M., Elad, M., Bruckstein, A., et al. (2006) K-svd: an algorithm for designing overcomplete dictionaries for sparse representation. *IEEE Transactions on Signal Processing*, 54(11), 4311.

Ahonen, T., Hadid, A., Pietikainen, M. (2006). Face description with local binary patterns: application to face recognition. *IEEE Transactions on Pattern Analysis & Machine Intelligence*, (12), 2037–2041.

Belhumeur, P.N., Hespanha, J.P., Kriegman, D.J. (1997). *Eigenfaces vs. Fisherfaces: Recognition Using Class Specific Linear Projection*. Technical report, Yale University, New Haven, United States of America.

Beymer, D. (1994). Face recognition under varying pose. In: *CVPR*, Vol. 94, page 137, Citeseer.

Blanz, V., Vetter, T. (2003). Face recognition based on fitting a 3d morphable model. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 25(9), 1063–1074.

Chai, X., Shan, S., Chen, X., Gao, W. (2007). Locally linear regression for pose-invariant face recognition. *IEEE Transactions on Image Processing*, 16(7), 1716–1725.

Cootes, T.F., Edwards, G.J., Taylor, C.J. (2001). Active appearance models. *IEEE Transactions on Pattern Analysis & Machine Intelligence*, (6), 681–685.

Ding, C., Tao, D. (2016). A comprehensive survey on pose-invariant face recognition. *ACM Transactions on Intelligent Systems and Technology (TIST)*, 7(3), 37.

Elad, M. (2010). From exact to approximate solutions. In: *Sparse and Redundant Representations*, Springer, pp. 79–109.

Engan, K., Aase, S.O., Husoy, J.H. (1999). Method of optimal directions for frame design. In: *1999 IEEE International Conference on Acoustics, Speech, and Signal Processing, 1999. Proceedings*, Vol. 5. IEEE, pp. 2443–2446.

Gottumukkal, R., Asari, V.K. (2004). An improved face recognition technique based on modular pca approach. *Pattern Recognition Letters*, 25(4), 429–436.

Gross, R., Matthews, I., Baker, S. (2004). Appearance-based face recognition and light-fields. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 26(4), 449–465.

Jiang, D., Hu, Y., Yan, S., Zhang, L., Zhang, H., Gao, W. (2005). Efficient 3d reconstruction for face recognition. *Pattern Recognition*, 38(6), 787–798.

Jiang, Z., Lin, Z., Davis, L.S. (2013). Label consistent k-SVD: learning a discriminative dictionary for recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 35(11), 2651–2664.

Lee, H.-S., Kim, D. (2006). Generating frontal view face image for pose invariant face recognition. *Pattern Recognition Letters*, 27(7), 747–754.

Liu, X., Chen, T. (2005). Pose-robust face recognition using geometry assisted probabilistic modeling. In: *IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2005. CVPR 2005*, Vol. 1. IEEE, pp. 502–509.

Mairal, J., Ponce, J., Sapiro, G., Zisserman, A., Bach, F.R. (2009). Supervised dictionary learning. *Advances in Neural Information Processing Systems*, 1033–1040.

McKenna, S.J., Gong, S., Collins, J.J. (1996). Face tracking and pose representation. In: *BMVC*. Citeseer, pp. 1–10.

Messer, K., Kittler, J., Sadeghi, M., Hamouz, M., Kostin, A., Cardinaux, F., Marcel, S., Bengio, S., Sanderson, C., Poh, N., et al. (2004). Face authentication test on the banca database. In: *Proceedings of the 17th International Conference on Pattern Recognition, 2004. ICPR 2004*, Vol. 4. IEEE, pp. 523–532.

Murase, H., Nayar, S.K. (1993). Learning and recognition of 3d objects from appearance. In: *IEEE Qualitative Vision Workshop*. CVPR, New York.

Murphy-Chutorian, E., Trivedi, M.M. (2009). Head pose estimation in computer vision: a survey. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 31(4), 607–626.

Nastar, C., Mitschke, M. (1998). Real-time face recognition using feature combination. In: *Third IEEE International Conference on Automatic Face and Gesture Recognition, 1998. Proceedings*. IEEE, pp. 312–317.

Pentl, A., Moghaddam, B., Starner, T. (1994). View-based and modular eigenspaces for face recognition. In: *Proceedings/CVPR, IEEE Computer Society Conference on Computer Vision and Pattern Recognition. IEEE Computer Society Conference on Computer Vision and Pattern Recognition*.

Phillips, P.J., Moon, H., Rizvi, S.A., Rauss, P.J. (2000). The feret evaluation methodology for face-recognition algorithms. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(10), 1090–1104.

Prince, S.J.D., Elder, J.H., Warrell, J., Felisberti, F.M. (2008). Tied factor analysis for face recognition across large pose differences. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 30(6), 970–984.

Sharma, A., Al Haj, M., Choi, J., Davis, L.S., Jacobs, D.W. (2012). Robust pose invariant face recognition using coupled latent space discriminant analysis. *Computer Vision and Image Understanding*, 116(11), 1095–1110.

Sim, T., Baker, S., Bsat, M. (2002). The cmu pose, illumination, and expression (pie) database. In: *Fifth IEEE International Conference on Automatic Face and Gesture Recognition, 2002. Proceedings*. IEEE, pp. 53–58.

Wright, J., Ma, Y., Mairal, J., Sapiro, G., Huang, T.S., Yan, S. (2010). Sparse representation for computer vision and pattern recognition. *Proceedings of the IEEE*, 98(6),1031–1044.

Wright, J., Yang, A.Y., Ganesh, A., Shankar Sastry, S., Ma, Y. (2009). Robust face recognition via sparse representation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 31(2), 210–227.

Yu, H., Liu, H. (2014). Facial pose estimation via dense and sparse representation. In: *2014 IEEE Symposium on Robotic Intelligence In Informationally Structured Space (RiiSS)*. IEEE, pp. 1–6.

Zhang, H., Zhang, Y., Huang, T.S. (2013). Pose-robust face recognition via sparse representation. *Pattern Recognition*, 46(5),1511–1521.

Zhang, X., Gao, Y. (2009). Face recognition across pose: a review. *Pattern Recognition*, 42(11), 2876–2896.

Zhang, X., Gao, Y., Leung, M.K.H. (2008). Recognizing rotated faces from frontal and side views: an approach toward effective use of mugshot databases. *IEEE Transactions on Information Forensics and Security*, 3(4), 684–697.

Zhang, X., Pham, D.-S., Venkatesh, S., Liu, W., Phung, D. (2015). Mixed-norm sparse representation for multi view face recognition. *Pattern Recognition*, 48(9), 2935–2946.

Zhao, L., Zhang, Y., Yin, B., Sun, Y., Hu, Y., Piao, X., Wu, Q. (2016). Fisher discrimination-based l2,1-norm sparse representation for face recognition. *The Visual Computer*, 32(9), 1165–1178.

Zhou, Z., Fu, J.H., Zhang, H., Chen, Z. (2001). Neural network ensemble based view invariant face recognition. *Journal of Computer Study and Development*, 38(9), 1061–1065.

A. Farahani received his BSc from Arak University, Arak, Iran, in software engineering in 2014. Then he pursued his MSc in artificial intelligence in Shahid Bahonar University of Kerman, Iran, and received his master degree in 2017 under the supervision of Dr Hadis Mohseni. His research interests include pattern recognition, supervised and unsupervised learning methods and their applications in image processing.

H. Mohseni received his BSc in hardware engineering from Sharif University of Technology (SUT), Tehran, Iran, in 2004. Then she continued her MSc in artificial intelligence in SUT and received her degree in 2007 working on medical image processing. She then pursued her PhD in artificial intelligence in SUT working on multi-pose face recognition and received her PhD degree in 2013 under the supervision of Prof. Shohreh Kasaei. Now she is an assistant professor in Shahid Bahonar University of Kerman and her research interests include pattern recognition, image and video processing, medical image processing and deep learning.