Dictionary-Based Face Recognition Under Variable Lighting and Pose

Vishal M. Patel, *Member, IEEE*, Tao Wu, *Student Member, IEEE*, Soma Biswas, *Member, IEEE*, P. Jonathon Phillips, *Fellow, IEEE*, and Rama Chellappa, *Fellow, IEEE*

***Abstract*—We present a face recognition algorithm based on simultaneous sparse approximations under varying illumination and pose. A dictionary is learned for each class based on given training examples which minimizes the representation error with a sparseness constraint. A novel test image is projected onto the span of the atoms in each learned dictionary. The resulting residual vectors are then used for classification. To handle variations in lighting conditions and pose, an image relighting technique based on pose-robust albedo estimation is used to generate multiple frontal images of the same person with variable lighting. As a result, the proposed algorithm has the ability to recognize human faces with high accuracy even when only a single or a very few images per person are provided for training. The efficiency of the proposed method is demonstrated using publicly available databases and it is shown that this method is efficient and can perform significantly better than many competitive face recognition algorithms.**

***Index Terms*—Biometrics, dictionary learning, face recognition, illumination variation, outlier rejection.**

I. Introduction

FACE recognition is a challenging problem that has been actively researched for over two decades [1]. Current systems work very well when the test image is captured under controlled conditions. However, their performance degrades significantly when the test image contains variations that are not present in the training images. Some of these variations include illumination, pose, expression, cosmetics, and aging.

It has been observed that since human faces have similar overall configuration, face images can be described by a relatively low dimensional subspace. Dimensionality reduction subspace methods such as principle component analysis (PCA) [2], linear discriminant analysis (LDA) [3], [4] and independent component analysis (ICA) [5] have been proposed for the task of face recognition. These approaches can be classified into either generative or discriminative methods. One of the major advantages of using generative approaches is that they are known to be less sensitive to noise than the discriminative approaches [1].

In recent years, theories of sparse representation (SR) and compressed sensing (CS) have emerged as powerful tools for efficiently processing data in nontraditional ways. This has led to a resurgence in interest in the principles of SR and CS for face recognition [6]–[11]. Wright *et al.* [7] introduced an algorithm, called sparse representation-based classification (SRC), where the training face images are the dictionary and a novel test image is classified by finding its sparse representation with respect to this dictionary. This work was later extended to handle misalignment and illumination variations [8], [9]. Also, Nagesh and Li presented an expression-invariant face recognition method using distributed compressed sensing and joint sparsity models in [10]. A face recognition method based on sparse representation for recognizing 3-D face meshes under expressions using low-level geometric features was presented by Li *et al.* in [11]. Phillips [6] proposed matching pursuit filters for face feature detection and identification. The filters were designed through a simultaneous decomposition of a training set into a 2-D wavelet expansion designed to discriminate among faces. It was shown that the resulting algorithm was robust to facial expression and the surrounding environment.

There are a number of hurdles that face recognition systems based on sparse representation must overcome. One is designing algorithms that are robust to changes in illumination; a second is that algorithms need to efficiently scale as the number of people enrolled in the system increases. In some of the previous approaches, the challenges mentioned are met by collecting a set of images of each person that spans the space of expected variations in illumination. The SRC approach recognizes faces by solving an optimization problem over the set of images enrolled into the database. This solution trades robustness and size of the database against computational efficiency.

In this paper, we present an algorithm to perform face recognition across varying illumination and pose based on learning small sized class specific dictionaries.[1] Our method consists of two main stages. In the first stage, given training samples from each class, class specific dictionaries are trained with some fixed number of atoms.[2] In the second stage, a novel test face image is projected onto the span of the atoms in each learned dictionary. The residual vectors are then used for classification. Furthermore, assuming the Lambertian reflectance model for the

Manuscript received May 10, 2011; revised February 12, 2012; accepted February 16, 2012. Date of publication February 27, 2012; date of current version May 08, 2012. The work of V. M. Patel, T. Wu, and R. Chellappa was supported in part by a Cooperative Agreement 70NANB11H023 from NIST and a MURI grant from the Office of Naval Research under Grant N00014-1-08-0638. The work of P. J. Phillips was supported by the Federal Bureau of Investigation. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Fabio Scotti.

V. M. Patel, T. Wu, and R. Chellappa are with the Department of Electrical and Computer Engineering and the Center for Automation Research, UMIACS, University of Maryland, College Park, MD 20742 USA (e-mail: pvishalm@umiacs.umd.edu; taowu@umiacs.umd.edu; rama@umiacs.umd.edu).

S. Biswas is with the Department of Computer Science and Engineering, University of Notre Dame, Notre Dame, IN 46556 USA (e-mail: sbiswas@nd.edu).

P. J. Phillips is with the National Institute of Standards and Technology, Gaithersburg, MD 20899 USA (e-mail: jonathon@nist.gov).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TIFS.2012.2189205

[1]A preliminary version of this work appeared in [12].

[2]Elements of a dictionary are commonly referred to as atoms.

surface of a face, we integrate a relighting approach within our framework so that we can add many elements to gallery and robustness to illumination and pose changes can be realized. In this setting, as will become apparent, our method has the ability to recognize faces even when only a single or a few images are provided for training.

This paper is organized as follows: In Section II, the proposed dictionary-based face recognition algorithm is detailed. Section III discusses the relighting approach for generating new illuminated images from a single face image. Experimental results are presented in Sections IV and Section V concludes the paper with a brief summary and discussion.

II. Dictionary-Based Recognition

In this section, we present our proposed face recognition algorithm based on learning class specific dictionaries.

A. Learning Class Specific Reconstructive Dictionaries

In face recognition, given labeled training images, the objective is to identify the class of a novel probe face image. Suppose that we are given C distinct classes and a set of m training images per class. We identify an $l \times q$ grayscale image as an N-dimensional vector, $\mathbf{x}$, which can be obtained by stacking its columns, where $N = l \times q$. Let

$$\mathbf{B}_i = [\mathbf{x}_i^1, \ldots, \mathbf{x}_i^m] \in \mathbb{R}^{N \times m} \quad (1)$$

be an $N \times m$ matrix of training images corresponding to the ith class.

In face recognition, there are numerous techniques that exploit the structure of the matrix $\mathbf{B}_i$ [1]. Images of the same person can vary significantly due to the variations present during the data capture process. Hence, it is essential to develop a method that extracts the common internal structure of given images and neglects minor variations. To this end, we seek a dictionary $\mathbf{D}_i \in \mathbb{R}^{N \times K}$ that leads to the best representation for each member in $\mathbf{B}_i$, under strict sparsity constraints. One can obtain this by solving the following optimization problem:

$$(\hat{\mathbf{D}}_i, \hat{\mathbf{\Gamma}}_i) = \arg \min_{\mathbf{D}_i, \mathbf{\Gamma}_i} \|\mathbf{B}_i - \mathbf{D}_i \mathbf{\Gamma}_i\|_F^2 \text{ s.t. } \forall i \ \|\boldsymbol{\gamma}_i^k\|_0 \leq T_0 \quad (2)$$

where $\boldsymbol{\gamma}_i^k \in \mathbb{R}^K, k \in \{1, \ldots, m\}$ represents a column of $\mathbf{\Gamma}_i \in \mathbb{R}^{K \times m}, T_0$ is a sparsity parameter and the ℓ_0 sparsity measure $\| \cdot \|_0$ counts the number of nonzero elements in the representation. Here, $\|\mathbf{A}\|_F$ denotes the Frobenius norm. One of the simplest algorithms for finding such a dictionary is the K-SVD algorithm [13].

The K-SVD algorithm is an iterative method and it alternates between sparse-coding and dictionary update steps. First, a dictionary $\mathbf{D}_i$ with ℓ_2 normalized columns is initialized. For example, this can be done by randomly selecting face images from the gallery set. Then, the main iteration is composed of the following two stages.

1) *Sparse coding*: In this step, $\mathbf{D}_i$ is fixed and the following optimization problem is solved to compute the representation vector $\boldsymbol{\gamma}_i^k$ for each example $\mathbf{x}_i^k, k \in \{1, \ldots, m\}$, i.e.,

$$k = 1, \ldots, m, \quad \min_{\boldsymbol{\gamma}_i^k} \|\mathbf{x}_i^k - \mathbf{D}_i \boldsymbol{\gamma}_i^k\|_2^2 \text{ s.t. } \|\boldsymbol{\gamma}_i^k\|_0 \leq T_0.$$

Since the problem is NP-hard, approximate solutions are usually sought. Any standard technique [14] can be used but a greedy pursuit algorithm such as orthogonal matching pursuit [15], [16] is often employed due to its efficiency [17].

2) *Dictionary update*: In this stage, the dictionary update is performed atom-by-atom in an efficient way. It has been observed that the K-SVD algorithm converges in a few iterations.

B. Classification Based on Learned Dictionaries

Given C distinct classes and m training images per class, let $\mathbf{B}_i$ be as defined in (1) for $i = 1, \ldots, C$. For training, we first learn C class specific dictionaries, $\mathbf{D}_i$, to represent the training samples in each $\mathbf{B}_i$, with some sparsity level, using the K-SVD algorithm. Once the dictionaries have been learned for each class, given a test sample $\mathbf{y}$, we project it onto the span of the atoms in each $\mathbf{D}_i$ using the orthogonal projector

$$\mathbf{P}_i = \mathbf{D}_i \left(\mathbf{D}_i^T \mathbf{D}_i\right)^{-1} \mathbf{D}_i^T. \quad (3)$$

The approximation and residual vectors can then be calculated as

$$\hat{\mathbf{y}}_i = \mathbf{P}_i \mathbf{y} = \mathbf{D}_i \boldsymbol{\alpha}_i \quad (4)$$

and

$$\mathbf{r}_i(\mathbf{y}) = \mathbf{y} - \hat{\mathbf{y}}_i = (\mathbf{I} - \mathbf{P}_i)\mathbf{y} \quad (5)$$

respectively, where $\mathbf{I}$ is the identity matrix and

$$\boldsymbol{\alpha}_i = \left(\mathbf{D}_i^T \mathbf{D}_i\right)^{-1} \mathbf{D}_i^T \mathbf{y} \quad (6)$$

are the coefficients. Since the K-SVD algorithm finds the dictionary, $\mathbf{D}_i$, that leads to the best representation for each examples in $\mathbf{B}_i$, we suspect $\|\mathbf{r}_i(\mathbf{y})\|_2$ to be small if $\mathbf{y}$ were to belong to the ith class and large for the other classes. Based on this, we can classify $\mathbf{y}$ by assigning it to the class, $d \in \{1, \ldots, C\}$, that gives the lowest reconstruction error, $\|\mathbf{r}_i(\mathbf{y})\|_2$

$$\begin{aligned} d &= \text{identity}(\mathbf{y}) \\ &= \arg \min_i \|\mathbf{r}_i(\mathbf{y})\|_2. \end{aligned} \quad (7)$$

An example of how our algorithm works is illustrated in Fig. 1.

C. Dealing With Small Arbitrary Noise

An assumption underlying the treatment given previously is that the test vector $\mathbf{y}$ is free of error. In practice, $\mathbf{y}$ will often be contaminated by some small noise perturbations. Hence, we consider the following more general model for $\mathbf{y}$:

$$\mathbf{y} = \tilde{\mathbf{y}} + \mathbf{z} \quad (8)$$

where $\tilde{\mathbf{y}}$ and $\mathbf{z}$ are the underlying noise free image and random noise term, respectively. Recall that constructing an approximation $\hat{\tilde{\mathbf{y}}}$ to $\tilde{\mathbf{y}}$ as

$$\hat{\tilde{\mathbf{y}}}_i = \mathbf{D}_i \boldsymbol{\alpha}_i$$

requires an estimation of $\boldsymbol{\alpha}_i$. In the case of least squares approximation, $\boldsymbol{\alpha}_i$ are those that minimize the following error:

$$\hat{\boldsymbol{\alpha}}_i = \min_{\boldsymbol{\alpha}_i} \|\mathbf{y} - \mathbf{D}\boldsymbol{\alpha}_i\|_2^2.$$

In this case, $\boldsymbol{\alpha}_i$ are given by (6). However, it is commonly known that the least squares method is sensitive to gross errors or out-

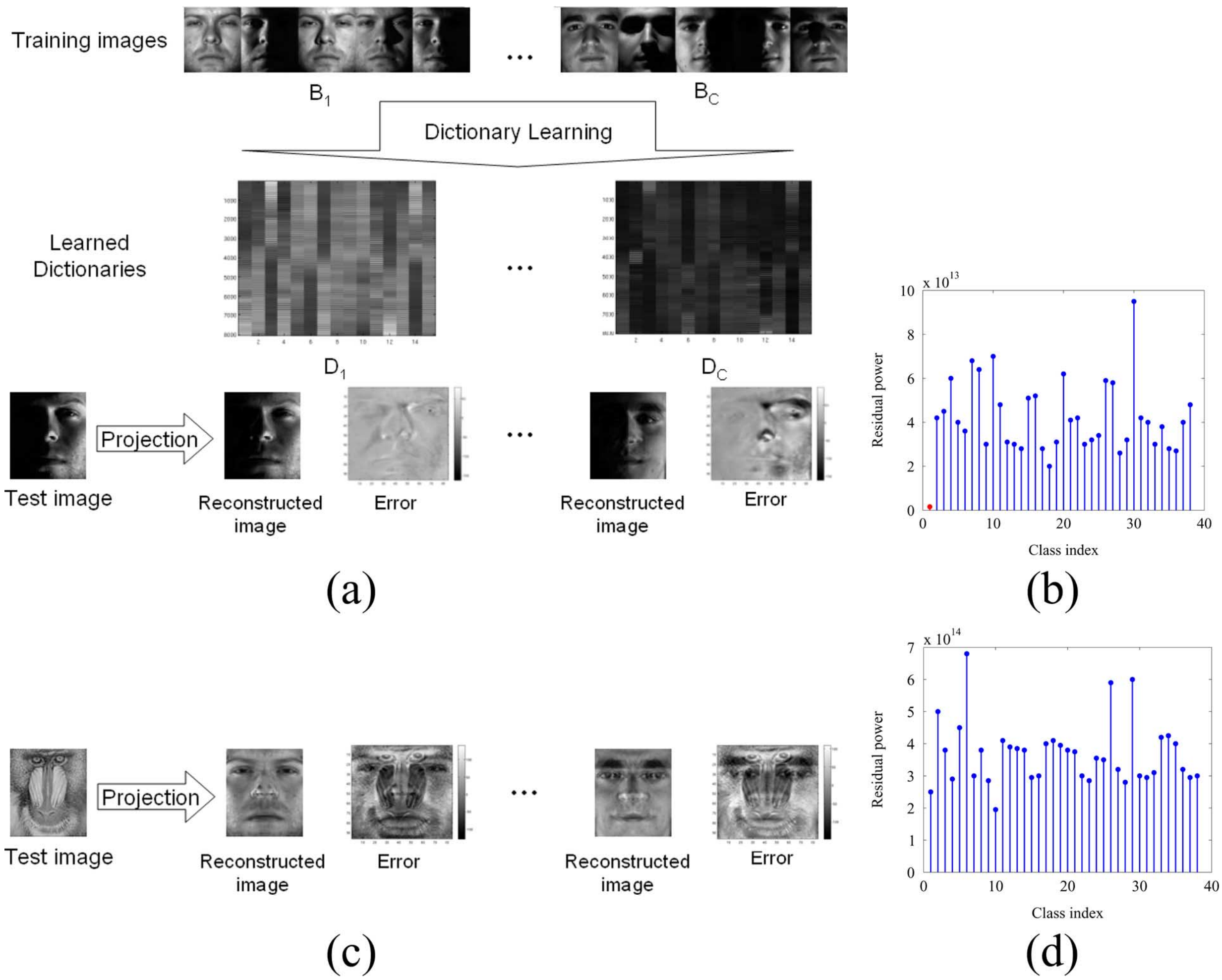


Fig. 1. Overview of our approach. (a) Given C sets of training images corresponding to C different faces, K-SVD algorithm is used to learn face specific dictionaries. Then, a novel test image is projected onto the span of the atoms in each of the learned dictionaries and the approximation errors are computed. (b) The class that is associated to a test image is then declared as the one that produces the smallest approximation error. In this example, class 1 is declared as the true class. (c) and (d) Example of a nonface test image and the resulting residuals, respectively.

liers. Hence, we need a formulation that recovers $\boldsymbol{\alpha}_i$ from the noisy observations (8) in a robust way. To robustly estimate the coefficients $\boldsymbol{\alpha}_i$ one can replace the quadratic error norm with a more robust error norm. This can be done by minimizing the following problem:

$$\hat{\boldsymbol{\alpha}}_i = \min_{\boldsymbol{\alpha}_i} \|\mathbf{y} - \mathbf{D}\boldsymbol{\alpha}_i\|_1$$

where $\| \mathbf{x} \|_1 = \sum_i |(x_i)|$. The resulting estimate is known as least absolute deviation (LAD) [18] and can be solved by linear programming methods.

D. Rejection Rule for Nonface Images

For classification, it is important to be able to detect and then reject invalid test samples. To decide whether a given test sample is valid or not, we define the following rejection rule.

Given a test image $\mathbf{y}$, for all classes in the training set, the score s_{yi} of the test image $\mathbf{y}$ to the ith class is computed as

$$s_{yi} = \frac{1}{\|\mathbf{r}_i(\mathbf{y})\|_2^2}$$

where $\mathbf{r}_i(\mathbf{y})$ is the residual vector as defined in (5). Then, for each test image $\mathbf{y}$, the score values are sorted in the decreasing order such that $s'_{y1} \geq s'_{y2} \geq \cdots \geq s'_{yC}$. The corresponding sorted classes are the candidate classes for each test image. The first candidate class is the most likely class that the test image belongs to. We define the ratio between the score of the first candidate class to the score of the second candidate class

$$\lambda_y = \frac{s'_{y1}}{s'_{y2}} \tag{9}$$

as a measure of the reliability of the recognition rate. Based on this, a threshold τ can be chosen such that $\mathbf{y}$ is accepted as a valid/good image if $\lambda_y \geq \tau$, otherwise rejected as an invalid/bad image. Since the score values to all the candidate classes are sorted, the score values of the third and the higher order candidates are less than or equal to the score of the second candidate class. Hence, a high ratio λ_y for the test image $\mathbf{y}$ would show that the score of the first candidate class is significantly greater than all the other scores. Therefore, the identification result can be claimed to be reliable.

To illustrate how this rejection rule works, consider the test images shown in Fig. 1(a) and (c). Since, the image shown in Fig. 1(a) belongs to class 1, the corresponding ratio comes out to be $\lambda_y = 26.29$, whereas the ratio corresponding to an invalid

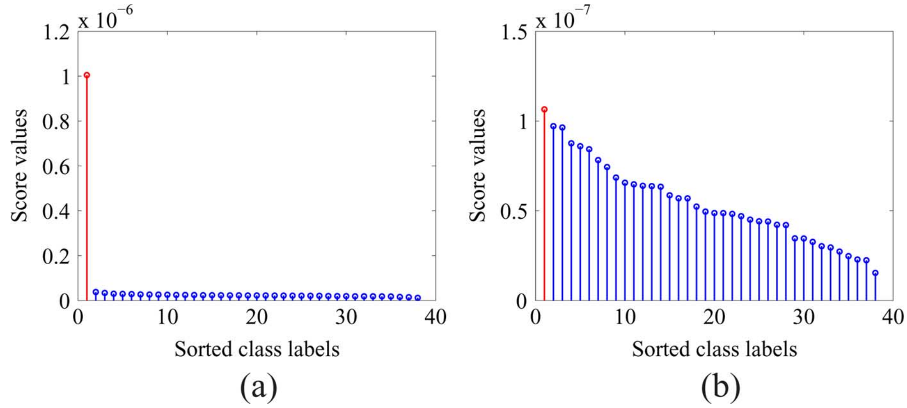


Fig. 2. Score values normalized using (9) and sorted. Plot (a) corresponds to the test image shown in Fig. 1(a) and plot (b) corresponds to the nonface test image shown in Fig. 1(c).

test image shown in Fig. 1(c) comes out to be $\lambda_y = 1.17$. Hence, setting a threshold, τ, high enough this nonface image can be rejected. This is illustrated in Fig. 2.

III. Face Recognition Across Varying Illumination and Pose

Images of the same person can vary significantly due to variations in illumination conditions. Hence, the performance of most existing face recognition algorithms is highly sensitive to illumination variations. In this section, we introduce a relighting method to deal with this illumination problem. The idea is to capture the illumination conditions that might occur in the test sample in the training samples.

A. Albedo Estimation

Assuming the Lambertian reflectance model for the facial surface, one can relate the surface normals, albedo and the intensity image by an image formation model. The diffused component of the surface reflection is given by

$$x_{i,j} = \rho_{i,j} \max\left(\mathbf{n}_{i,j}^T \mathbf{s}, 0\right) \tag{10}$$

where $x_{i,j}$ is the pixel intensity at position (i,j), $\mathbf{s}$ is the light source direction, $\rho_{i,j}$ is the surface albedo at position (i,j), $\mathbf{n}_{i,j}$ is the surface normal of the corresponding surface point and $1 \leq i \leq l, 1 \leq j \leq q$. The max function in (10) accounts for the formation of attached shadows. Neglecting the attached shadows, (10) can be linearized as

$$\begin{aligned} x_{i,j} &= \rho_{i,j} \max\left(\mathbf{n}_{i,j}^T \mathbf{s}, 0\right) \\ &\approx \rho_{i,j} \mathbf{n}_{i,j}^T \mathbf{s}. \end{aligned} \tag{11}$$

Let $\mathbf{n}_{i,j}^{(0)}$ and $\mathbf{s}^{(0)}$ be the initial values of the surface normal and illumination direction. These initial values can be domain dependent average values. The Lambertian assumption imposes the following constraints on the initial albedo:

$$\rho_{i,j}^{(0)} = \frac{x_{i,j}}{\mathbf{n}_{i,j}^{(0)} \cdot \mathbf{s}^{(0)}} \tag{12}$$

where $\cdot$ denotes the standard dot product operation. Using (11), (12) can be rewritten as

$$\begin{aligned} \rho_{i,j}^{(0)} &= \rho_{i,j} \frac{\mathbf{n}_{i,j} \cdot \mathbf{s}}{\mathbf{n}_{i,j}^{(0)} \cdot \mathbf{s}^{(0)}} = \rho_{i,j} + \frac{\mathbf{n}_{i,j} \cdot \mathbf{s} - \mathbf{n}_{i,j}^{(0)} \cdot \mathbf{s}^{(0)}}{\mathbf{n}_{i,j}^{(0)} \cdot \mathbf{s}^{(0)}} \rho_{i,j} \\ &= \rho_{i,j} + \omega_{i,j} \end{aligned} \tag{13}$$

where

$$\omega_{i,j} = \frac{\mathbf{n}_{i,j} \cdot \mathbf{s} - \mathbf{n}_{i,j}^{(0)} \cdot \mathbf{s}^{(0)}}{\mathbf{n}_{i,j}^{(0)} \cdot \mathbf{s}^{(0)}} \rho_{i,j}.$$

This can be viewed as a signal estimation problem where ρ is the original signal, $\rho^{(0)}$ is the degraded signal and ω is the signal dependent noise. Using this model, the albedo can be estimated using the method of minimum mean squared error criterion [19]. Then, using the estimated albedo map, one can generate new images for a given light source direction using the image formation model in (10). This can be done by combining the estimated albedo map and light source direction with the average facial information [20].

B. Image Relighting

It has been found that the set of images under all possible illumination conditions can be well approximated by a 9-D linear subspace [21]. Based on this result, Lee *et al.* [21] showed that there exists a configuration of nine light source directions such that the subspace formed by the images taken under these nine sources is effective for recognizing faces under a wide range of lighting conditions. The nine prespecified light source directions are given by [21]

$$\begin{aligned} \phi &= \{0, 49, -68, 73, 77, -84, -84, 82, -50\}^\circ \\ \theta &= \{0, 17, 0, -18, 37, 47, -47, -56, -84\}^\circ. \end{aligned}$$

Hence, the image formation equation can be rewritten as

$$\mathbf{x} = \sum_{i=1}^{9} a_i \mathbf{x}_i \tag{14}$$

Fig. 3. Examples of original images (first column) and corresponding relighted images with different light source directions from PIE data set.

where $\mathbf{x}_i = \rho \max(\mathbf{n}^T \mathbf{s}_i, 0)$, and $\{\mathbf{s}_1, \ldots, \mathbf{s}_9\}$ are the prespecified illumination directions. To characterize the set of images under various illumination conditions, one can generate images under the nine prespecified illumination directions and use them in the gallery. By generating multiple face images with different lighting from a single face image, one can achieve good recognition accuracy even when only a single or a very few images are provided for training. Fig. 3 shows some relighted images and the corresponding input images.

C. Pose-Robust Albedo Estimation

The method presented previously can be generalized such that it can handle pose variations [22]. Let $\bar{\mathbf{n}}_{i,j}$, $\bar{\mathbf{s}}$ and $\bar{\Theta}$ be some initial estimates of the surface normals, illumination direction and initial estimate of surface normals in pose Θ, respectively. Then, the initial albedo at pixel (i, j) can be obtained by

$$\bar{\rho}_{i,j} = \frac{x_{i,j}}{\bar{\mathbf{n}}_{i,j}^{\bar{\Theta}} \cdot \bar{\mathbf{s}}}$$

where $\bar{\mathbf{n}}_{i,j}^{\bar{\Theta}}$ denotes the initial estimate of surface normals in pose $\bar{\Theta}$. Using this model, we can reformulate the problem of recovering albedo as a signal estimation problem. Using arguments similar to (12), we get the following formulation for the albedo estimation problem in the presence of pose:

$$\bar{\rho}_{i,j} = \rho_{i,j} h_{i,j} + \omega_{i,j}$$

where

$$w_{i,j} = \frac{\bar{\mathbf{n}}_{i,j}^{\Theta} \cdot \mathbf{s} - \bar{\mathbf{n}}_{i,j}^{\Theta} \cdot \bar{\mathbf{s}}}{\bar{\mathbf{n}}_{i,j}^{\bar{\Theta}} \cdot \bar{\mathbf{s}}} \rho_{i,j}$$
$$h_{i,j} = \frac{\bar{\mathbf{n}}_{i,j}^{\Theta} \cdot \bar{\mathbf{s}}}{\bar{\mathbf{n}}_{i,j}^{\bar{\Theta}} \cdot \bar{\mathbf{s}}}$$

$\rho_{i,j}$ is the true albedo and $\bar{\rho}_{i,j}$ is the degraded albedo. In the case when the pose is known accurately, $\bar{\Theta} = \Theta$ and $h_{i,j} = 1$. Hence, this can be viewed as a generalization of (13) in the case of unknown pose. Using this model, a stochastic filtering framework was recently presented in [22] to estimate the albedo from a single nonfrontal face image. Once pose and illumination have been normalized, one can use the relighting method described in the previous section to generate multiple frontal images with different lighting to achieve illumination and pose-robust recognition. Fig. 4 shows some examples of pose normalized images using this method.

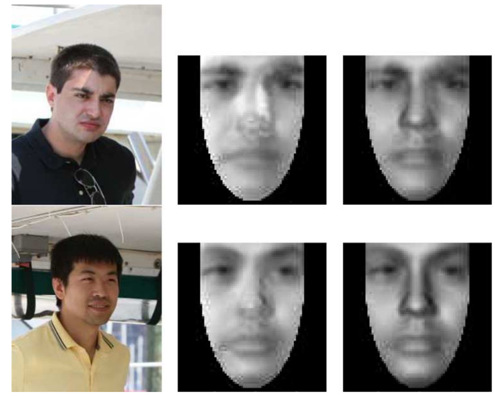


Fig. 4. Pose-robust albedo estimation. Left column: Original input images. Middle column: Recovered albedo maps corresponding to frontal face images. Right column: Pose normalized relighted images.

Given a test sample $\mathbf{y}$ and C training matrices $\mathbf{B}_1, \cdots, \mathbf{B}_C$ where each $\mathbf{B}_i \in \mathbb{R}^{N \times m}$ contains m training samples.

Procedure:

1. For each training image, use the relighting approach described in section III to generate multiple images with different illumination conditions and use them in the gallery.
2. Learn the best dictionaries $\mathbf{D}_i$, to represent the face images in $\mathbf{B}_i$, using the K-SVD algorithm.
3. Compute the approximation vectors, $\hat{\mathbf{y}}_i$, and the residual vectors, $\mathbf{r}_i(\mathbf{y})$, using (4) and (5), respectively for $i = 1, \cdots, C$.
4. Identify $\mathbf{y}$ using (7).

Fig. 5. DFR algorithm.

We summarize our dictionary-based face recognition (DFR) algorithm in Fig. 5. Note that a K-SVD-based face recognition algorithm was recently proposed in [23], but we differ from this work in a few key areas. Unlike [23], we do not take discriminative approach to face recognition. Our method is a reconstructive approach to discrimination and does not require multiple images to be available. Another difference is that our algorithm has the ability to identify and reject nonface images.

IV. Experimental Results

To illustrate the effectiveness of our method, we present experimental results on three available databases for face recognition such as the Extended Yale B dataset [24], the AR dataset [25] and PIE dataset [26]. We also present the experimental results on a remote face database which has been acquired in an unconstrained outdoor maritime environment [27]. In the experiments with the Extended Yale B, PIE, and AR face datasets, the input face and eye locations are detected automatically using the Viola-Jones object detection framework [28]. The cropped faces are then aligned using the center of the eyes located by the

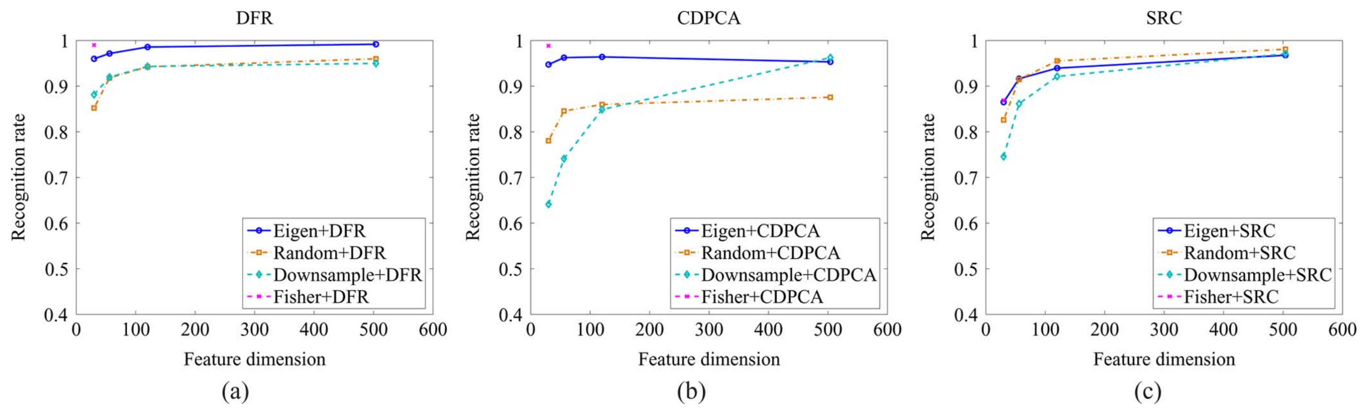


Fig. 6. Performance comparison on Extended Yale B database with various features, feature dimensions and methods. (a) Our method (DFR), (b) CDPCA, and (c) SRC [7].

Viola-Jones algorithm. An implementation of this object detection framework can be found in the OpenCV library [29].

The comparison with other existing face recognition methods in [7] suggests that the SRC algorithm is among the best. Hence, we treat it as state-of-the-art and use it as a bench mark for comparisons in this paper. The methods compared in [7] include nearest neighbor (NN), nearest subspace (NS), support vector machines (SVM) [30].

In all of our experiments, the K-SVD [13] algorithm is used to train the dictionaries with 15 atoms unless otherwise stated. The performance of our algorithm is compared with that of SRC and class dependent principal component analysis (CDPCA) [31]. Our algorithm is also tested using several features, namely, Eigenfaces, Fisherfaces, Randomfaces, and downsampled images.[3] All the experiments were done on a Linux system with Intel Xeon E5506/2.13 GHz processor using Matlab.

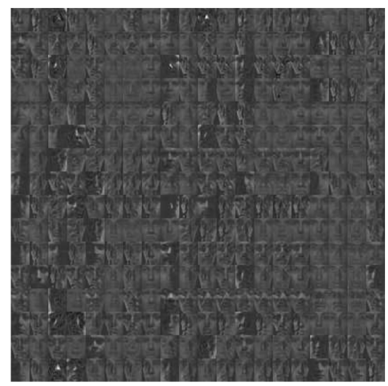


Fig. 7. A few learned dictionaries from the Extended Yale B dataset. Each row corresponds to a learned dictionary. By looking at each row, we see that the learned atoms are able to extract the common internal structure of images belonging to the same class and are able to remove much of the illumination.

A. Results on Extended Yale B Database

There are a total of 2 414 frontal face images of 38 individuals in the Extended Yale B database. These images were captured under various controlled indoor lighting conditions. They were cropped and normalized to the size of 192×168 [21].

Our first set of experiments on the Extended Yale B data set consists of testing the performance of our algorithm with different features and dimensions. The objective is to verify the ability of our algorithm in recognizing faces with different illumination conditions. We follow the experimental setup as considered in [7]. The feature space dimensions of 30, 56, 120, and 504 corresponding to the downsampling ratios of, $1/32, 1/24, 1/16$, and $1/8$, respectively, are computed. We randomly select 32 images per subject (i.e., half of the images) for training and the other half for testing. Recognition rates of different methods with different dimensions and features are compared in Fig. 6.

The maximum recognition rates achieved by DFR are 95.99%, 97.16%, 98.58%, and 99.17% for all 30, 56, 120 and 504 dimensional feature spaces, respectively. The maximum recognition rate achieved by SRC is 98.1% with 504 random faces [7]. Also, NN, NS, and SVM achieve the maximum recognition rates of 90.7%, 94.1%, and 97.7%, respectively [7]. CDPCA also performed quite well on this experiment. It achieved the maximum recognition rate of 96.24%. As can be seen from Fig. 6, the DFR performs favorably over some of the competitive methods for face recognition on the Extended Yale B database.

In Fig. 7, we show some of the learned dictionaries from the Extended Yale B dataset. In Fig. 7, each row corresponds to a learned dictionary. By looking at each row, we see that the learned atoms are able to extract the common internal structure of images belonging the same class and are able to remove much of the illumination.

B. Results on PIE Database

The PIE database contains face images of 68 subjects. The images were captured under 13 different poses and 21 flashes under pose, illumination and expression variations. The face images were cropped with the size 48×40. In the first set of experiments on the PIE data set, our objective is to perform recognition across illumination with images from one illumination condition forming the gallery while images from another illumination condition forming the test set. In this setting, there is just one image per subject in each gallery and probe set. See [32] for more details on how the training and test data sets are

[3]This means that, we first transform the given images into a feature space. We then train dictionaries on the feature space.

TABLE I

RECOGNITION RESULT ON FRONTAL IMAGES IN PIE DATA SET. f_i DENOTES IMAGES WITH iTH FLASH ON AS LABELED IN PIE DATA SET. EACH (i, j)TH ENTRY IS RANK-1 RECOGNITION RATE OBTAINED WITH IMAGES IN f_i AS GALLERY AND f_j AS PROBE SETS

Probe	f_8	f_9	f_{11}	f_{12}	f_{13}	f_{14}	f_{15}	f_{16}	f_{17}	f_{20}	f_{21}	f_{22}	Avg	Avg:MA[32]	Avg:MB[32]	Avg [19]
Gallery																
f_8	-	100	100	100	99	99	97	94	79	99	97	97	96	89	92	87
f_9	100	-	100	100	100	99	99	99	97	99	97	97	99	93	97	95
f_{11}	100	100	-	100	100	100	99	99	96	100	99	99	99	92	95	92
f_{12}	100	100	100	-	100	100	100	99	99	100	100	100	100	96	98	98
f_{13}	100	100	100	100	-	100	100	100	100	100	100	99	100	98	100	99
f_{14}	100	99	100	100	100	-	100	100	100	100	100	100	100	99	99	98
f_{15}	96	97	99	100	100	100	-	100	100	100	100	99	99	96	97	96
f_{16}	96	97	100	100	96	99	99	-	100	97	100	100	98	91	94	93
f_{17}	78	85	94	99	97	96	99	99	-	90	96	96	93	80	87	83
f_{20}	100	100	100	100	99	99	99	99	97	-	100	100	99	91	95	93
f_{21}	97	100	100	100	100	100	100	100	99	100	-	100	100	96	99	97
f_{22}	97	97	100	99	100	100	100	100	100	100	100	-	99	98	98	97
Avg	97	98	99	100	99	99	99	99	97	99	99	99	**99**	-	-	-
Avg [32]	88	94	93	97	99	99	96	89	75	93	98	98	-	93	-	-
Avg [32]	90	97	94	99	99	99	98	93	87	95	99	99	-	-	96	-
Avg [19]	91	97	93	99	99	98	94	91	80	93	99	96	-	-	-	94

TABLE II

RANK-1 RECOGNITION RESULTS (IN %) ON PIE DATASET

c_{05}	f_{21}	f_{20}	f_{12}	f_{11}	f_9	f_8
DFR	97	97	100	97	100	97
[34]	96	94	99	98	96	93
[22]	96	97	99	97	97	97
[33]	98	97	98	97	97	94

TABLE III

RANK-1 RECOGNITION RESULTS (IN %) ON PIE DATASET WITH DIFFERENT POSES AND ILLUMINATION VARIATIONS

	f_{21}	f_{20}	f_{12}	f_{11}	f_9	f_8
c_{07}	94	93	93	88	86	87
c_{09}	94	92	97	94	97	99
c_{29}	92	96	92	96	93	96
c_{37}	61	70	68	75	64	76

created for this experiment. The rank-1 results obtained using our method are reported in Table I. As can be seen from Table I, that our method achieves recognition rate over 99% in most of the experiments and on average it achieves the recognition rate of 99%.

For comparison, we have also included the average recognition rates from [32] and [19] which follow a similar experimental setting. In Table I, MA and MB correspond to method A and method B as presented in [32]. Based on their albedo estimation method, Biswas *et al.* [19] report an average recognition rate of 94%. Our results on this data set are also comparable to that of [33] and [34]. Using only f_{12} as the gallery set, [33] reported an average recognition rate of 98% for color images. Similarly, an average recognition rate of 99% (with f_{12} as gallery) is reported by Zhang and Samaras using their spherical harmonics based approach [34]. Based on gray scale images, we obtain an average recognition rate of 100% when f_{12} is used as the gallery. Furthermore, DFR is much faster than the algorithms presented in [33] and [34] with an advantage that it can deal with images of much smaller size.

In the second set of experiments using this database, we test the ability of our algorithm in recognizing faces in the presence of different poses and illumination. In particular, we use the images corresponding to f_{12} as the gallery set which contains the images in frontal pose (camera 27) and frontal illumination. The probe images are in side pose (camera 5) with various illumination conditions. See [26] for more details on camera, c, and flash, f, positions corresponding to this dataset. Each gallery and probe set contains just one image per subject. Table II reports the rank-1 recognition rates achieved by different methods. It can be seen that the proposed dictionary-based method performs favorably with some of the competitive methods [22], [34], [33].

To better analyze the robustness of our method to pose variations, we repeat the previous experiment on the PIE dataset with different poses. In particular, we select four poses corresponding to cameras 07, 09, 29, and 37 with different illumination conditions as the probe set. Table III reports the rank-1 recognition rates achieved by our method. As can be seen from this table, even in the presence of extreme pose variation (camera 37) our method is able to provide reasonable recognition performance.

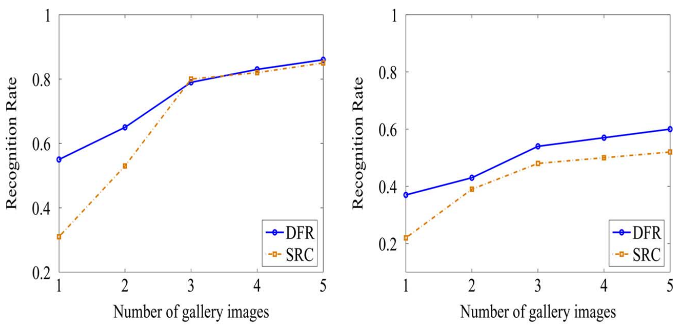
C. Results on AR Database

The AR database consists of over 4 00 frontal face images of 126 subjects (70 men and 56 women). All the images were converted to gray scale and cropped with the size of 165×120. The images feature frontal view faces with different facial expression, illumination variation and occlusion. Hence, this database is more challenging than the Yale B and PIE datasets. In this experiment, we choose a subset of the images consisting of 50 male subjects and 50 female subjects. 14 images per subject with illumination variations and expressions are used. From these 14 images, seven images from Session 1 are used for training and the other seven from Session 2 are used for testing [7].

The best recognition rate achieved by our algorithm is 93.7% which is a little lower than that of SRC and SVM whose reported best recognition rates are 94.7% and 95.7%, respectively [7]. NN and NS achieve the recognition rates of 89.7% and 90.3%, respectively [7] whereas CDPCA achieves the recognition rate of 59.00%.

(a) (b)

Fig. 8. A few cropped face images from remote face dataset. (a) Sample images from illumination folder. (b) Sample images from pose folder.



(a) (b)

Fig. 9. Recognition results on remote face dataset corresponding to (a) illumination folder and (b) pose folder.

D. Experiment on a Remote Face Dataset

In this section, we evaluate the effectiveness of our method on a remote face dataset [27]. In this dataset, a significant number of images are taken from long distances and under unconstrained outdoor environments. The distance from which the face images are taken varies from 5 to 250 m under different scenarios. Since all the faces in the data set could not be extracted reliably using existing state-of-the-art face detection algorithms and the faces only occupied small regions in large background scenes, the faces were manually cropped and rescaled to a fixed size [27]. The database contains 17 different individuals and 2102 face images in total. The number of faces per subject varies from 29 to 306. All the images are 120×120 pixel png images. The images are partitioned into various folders corresponding to different variations present during the data capture process. We only use the folders containing images with illumination and pose variations. Five clear images from each class are used for training and the rest of the images from the corresponding folders are used as the test set. Sample images from the illumination and pose folders are shown in Fig. 8.

The number of images in gallery is varied from one to five images per subject. The rank-1 recognition results obtained using SRC and DFR are compared in Fig. 9. The best recognition rate achieved by DFR on the images containing illumination is 85.8% compared to 85% for the SRC method. On the pose folder, DFR significantly outperforms the SRC method. The best recognition rate achieved by the SRC method on the pose folder is 52% whereas the DFR method achieves 60%.

E. Recognition With Partial Face Features

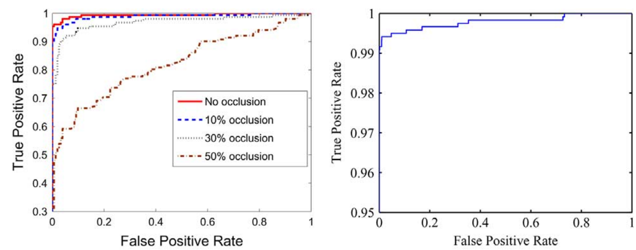
In this section, we report the ability of our algorithm in recognizing faces from the partial face features. Partial face features have been used in recovering the identity of human faces before [6], [7], [35]. We use the images in the Extended Yale B database for this experiment. For each subject, 32 images are randomly selected for training, and the remaining images are used for testing. The region of eye, nose and mouth are selected as partial face features [7]. These partial facial parts are manually cropped. Examples of these features are shown in Fig. 10. Note that in this experiment, we omit the relighting step of our algorithm. We learn dictionaries directly on the partial facial features. Table IV compares the results obtained by using our method with the other methods presented in [7]. As can be seen from the table, our method achieves recognition rates of 99.3%, 98.8% and 99.8% on eye, nose and mouth region, respectively, and it significantly outperforms SRC, NN, NS and SVM [7].

(a) (b) (c)

Fig. 10. Examples of partial facial features. (a) Eye, (b) nose, and (c) mouth.

TABLE IV
RECOGNITION RESULTS WITH PARTIAL FACIAL FEATURES

	Right Eye	Nose	Mouth
Dimension	5,040	4,270	12,936
DFR	**99.3%**	**98.8%**	**99.8%**
SRC	93.7%	87.3%	98.3%
NN	68.8%	49.2%	72.7%
NS	78.6%	83.7%	94.4%
SVM	85.8%	70.8%	95.3%



(a) (b)

Fig. 11. (a) ROC curves corresponding to rejecting outliers. Solid curve is generated by the DFR method based on our rejection rule. Dotted curves correspond to cases when different levels of occlusion has been added to the test images. (b) ROC curve corresponding to rejecting invalid test samples.

F. Rejecting Nonface Images

In this section, we demonstrate the effectiveness of our method in dealing with invalid test images with and without block occlusion. We test our rejection rule, described in Section II-D, on the Extended Yale B data set. We use Subsets 1 and 2 for training and Subset 3 for testing. We simulate varying levels of occlusion by replacing a randomly chosen block of each test image with random noise. We include only half of the subjects in the training set. This way, half of the subjects in the test set are new to the algorithm. We plot the receiver operating characteristic (ROC) curves according to different τ values in Fig. 11(a). As can be seen from this figure, that simple rejection rule performs quite well. It performs nearly perfectly at 10% occlusion and without any occlusion. Even at 50% occlusion, it performs better than making a random decision.

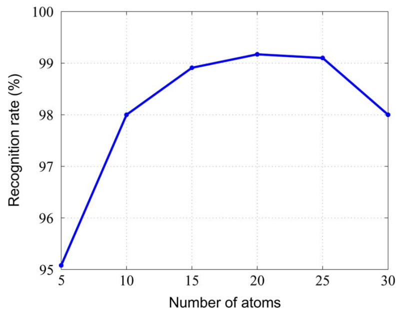


Fig. 12. Recognition rate versus number of dictionary atoms on Extended Yale B dataset.

This performance, can be further improved by applying our DFR method on different features such as PCA and LDA.

In the second set of experiments on rejecting invalid test samples, we use the same experimental set up as in Section IV-A. In order to test the ability of our rejection rule in rejecting invalid samples, we add 1198 randomly selected object images from the Eth80 dataset [36] to the probe set. The ROC curve corresponding to this experiment is shown in Fig. 11(b). As can be seen from this figure, that our simple rejection rule is able to remove most of the nonface images and performs nearly perfectly.

G. Recognition Rate Versus Number of Dictionary Atoms

In this section, we evaluate the performance of DFR as the number of trained dictionary atoms are changed. To this end, we repeat the experiment described in Section IV-A on DFR using 504 dimensional eigenfaces with different number of dictionary atoms. Fig. 12 shows the recognition rate versus number of atoms bar plot for this experiment. It can be observed that even selecting only five atoms per class dictionary, DFR provides a reasonable recognition performance on the Extended Yale B database. Experiments have shown that increasing the number of atoms more than 23 usually degrades the performance of our algorithm. This is the case because with more dictionary atoms the representation gets more exact and it has to deal with all the noise/distortion present in the data. Whereas with fewer number of dictionary atoms, much more accurate description of the internal structure of the class is captured and the robustness to distortions is realized [6], [37], [38].

H. Recognition Rate Versus Number of Training Images

In this section, we study the performance of DFR as we vary the number of training images in each class. We use the Extended Yale B database for the experiments in this section. All the images are scaled to the size of 64×64. We randomly select 1, 2, and 3 images per subject for training and the others for testing. We compare the performance of our method with that of SRC and dictionary-based SRC (DSRC).[4] For DSRC, we define

TABLE V
PERFORMANCE COMPARISON (IN %) OF DIFFERENT METHODS WITH RESPECT TO NUMBER OF TRAINING SAMPLES PER SUBJECT

No. images	DFR	SRC	DSRC	CDPCA	GF	LTV
1	**75.89**	32.37	30.98	5.52	66.65	67.92
2	**84.71**	37.20	44.23	26.22	76.35	79.61
3	**85.18**	37.45	53.57	30.25	77.18	84.93

a new training matrix $\mathbf{A}$ as the concatenation of learned dictionaries from all classes as

$$\mathbf{A} = [\mathbf{D}_1, \ldots, \mathbf{D}_C] \tag{15}$$

where $\mathbf{D}_i$ is the learned dictionary corresponding to the class matrix $\mathbf{B}_i$. Given a novel face image, $\mathbf{y}$, we solve the following ℓ_1-minimization problem to obtain the sparse coefficients

$$\hat{\beta} = \min_{\beta} \|\beta\|_1 \text{ subject to } \mathbf{A}\beta = \mathbf{y}. \tag{16}$$

Once the sparse solution is obtained, we follow the procedure of SRC to classify the test image. The experiment is carried out ten times and the average recognition rates of DFR along with SRC, DSRC and CDPCA are compared in Table V. This experiment shows that even in the presence of a few training images, our method can provide reasonable recognition of human faces. This performance can be further enhanced by learning dictionaries on features such as PCA and LDA.

Note that this experiment violates SRC's working premise that any test image that belongs to the same class will approximately lie in the linear span of the training samples from the corresponding class. As a result, SRC fails to provide reasonable recognition performance on this experiment. Similarly, 15 atoms per learned dictionary are not enough for DSRC to classify a novel test image with illumination variations via ℓ_1-minimization. Experiments have shown that the performance of DSRC generally increases as more number of atoms are kept in each learned dictionary.

We also compare the performance of our method with several state-of-the-art illumination normalization-based methods such as Gradient faces (GF) [39] and LTV [40]. Once the GF or LTV features are found, the recognition is performed by the nearest neighbor rule (in the case of 1 training image) or by the nearest subspace rule. The results are shown in the last two columns of Table V. As can be seen from Table V that DFR significantly outperforms GF and LTV based methods. Gradient faces and LTV features tend to be very noisy, especially in the dark regions of the face. This, in turn, effects the recognition performance. Since we are extending the gallery by adding multiple images of the subject with various illumination, DFR does not suffer from the above-mentioned artifacts and gives better recognition performance.

I. Efficiency

To illustrate the efficiency of our algorithm, in Table VI, we report the average runtime of DSRC and DFR in classifying a test sample with a gallery matrix containing 32 images from which 15 dictionary atoms are learned. As can be seen from the table that DFR is efficient even when the data dimension increases.

[4]Note that one can use the introduced relighting method to first enlarge the training set to capture the illumination variations and then use the SRC method for classification. However, as discussed earlier, with enlarged dictionary, the computational complexity of SRC increases tremendously. To reduce the complexity of the ℓ_1-minimization method, we first reduce the size of the enlarged dictionaries using small-sized learned dictionaries. We then apply the SRC method on a dictionary that is obtained by concatenating the learned dictionaries.

TABLE VI
AVERAGE RUNTIME IN SECONDS

Dimension	DFR	DSRC
30	$\mathbf{5.5 \times 10^{-4}}$	0.28
56	$\mathbf{6.8 \times 10^{-4}}$	0.63
120	$\mathbf{7.1 \times 10^{-4}}$	0.80
504	$\mathbf{1.5 \times 10^{-3}}$	0.97

J. Limitations

One limitation of the albedo estimation methods [19] and [22] is that they require the images to be aligned, as is the case with most state-of-the-art face recognition algorithms. The albedo estimation methods are also sensitive to facial expressions. Hence, when our method is used to estimate the albedo maps from a given face image with expressions, it produces artifacts in the final estimated albedo. As a result, DFR produces inferior recognition results on the databases with expressions such as AR face dataset.

V. Discussion and Conclusion

We have proposed a face recognition algorithm based on dictionary learning methods that is robust to changes in lighting and pose. This entails using a relighting approach based on a robust albedo estimation. Various experiments on popular face recognition data sets have shown that our method is efficient and can perform significantly better than many competitive face recognition algorithms.

Even though, in this paper, we took a reconstructive approach to dictionary learning, it is possible to learn discriminative dictionaries [37], [38], [41]–[48] for the task of face recognition, as was done in [23]. One of the main drawbacks of learning discriminative dictionaries is that it can tremendously increase the over all computational complexity which can make the real-time processing very difficult. Discriminative methods are also sensitive to noise. It remains an interesting topic for future work to develop and analyze the accuracy of a discriminative dictionary learning algorithm that is robust to pose, expression and illumination variations [49].

Acknowledgment

The authors would like to thank F. Scotti and the anonymous reviewers for their many valuable comments and suggestions which significantly improved this paper.

References

[1] W. Zhao, R. Chellappa, J. Phillips, and A. Rosenfeld, "Face recognition: A literature survey," *ACM Computing Surveys*, pp. 399–458, Dec. 2003.

[2] M. Turk and A. Pentland, "Eigenfaces for recognition," *J. Cognitive Neurosci.*, vol. 3, no. 1, pp. 71–86, 1991.

[3] P. Belhumeur, J. Hespanda, and D. Kriegman, "Eigenfaces versus fisherfaces: Recognition using class specific linear projection," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 19, no. 7, pp. 711–720, Jul. 1997.

[4] K. Etemad and R. Chellappa, "Discriminant analysis for recognition of human face images," *J. Opt. Soc. Amer. A*, vol. 14, no. 8, pp. 1724–1733, Aug. 1997.

[5] M. Bartlett, J. Movellan, and T. Sejnowski, "Face recognition by independent component analysis," *IEEE Trans. Neural Networks*, vol. 13, no. 6, pp. 1450–1464, Nov. 2002.

[6] P. J. Phillips, "Matching pursuit filters applied to face identification," *IEEE Trans. Image Process.*, vol. 7, no. 8, pp. 150–164, Aug. 1998.

[7] J. Wright, A. Y. Yang, A. Ganesh, S. S. Sastry, and Y. Ma, "Robust face recognition via sparse representation," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 31, no. 2, pp. 210–227, Feb. 2009.

[8] J. Huang, X. Huang, and D. Metaxas, "Simultaneous image transformation and sparse representation recovery," in *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, Ancorage, AK, Jun. 2008, pp. 1–8.

[9] A. Wagner, J. Wright, A. Ganesh, Z. Zhou, and Y. Ma, "Towards a practical face recognition system: Robust registration and illumination by sparse representation," in *Proc. IEEE Conf. Computer Vision Pattern Recognition*, Miami, FL, Jun. 2009, pp. 597–604.

[10] P. Nagesh and B. Li, "A compressive sensing approach for expression-invariant face recognition," in *Proc. IEEE Conf. Computer Vision Pattern Recognition*, Miami, FL, Jun. 2009, pp. 1518–1525.

[11] X. Li, T. Jia, and H. Zhang, "Expression-insensitive 3D face recognition using sparse representation," in *Proc. IEEE Conf. Computer Vision Pattern Recognition*, Miami, FL, Jun. 2009, pp. 2575–2582.

[12] V. M. Patel, T. Wu, S. Biswas, P. J. Phillips, and R. Chellappa, "Illumination robust dictionary-based face recognition," in *IEEE Int. Conf. Image Processing*, Brussels, Belgium, Oct. 2011.

[13] M. Aharon, M. Elad, and A. M. Bruckstein, "The K-SVD: An algorithm for designing of overcomplete dictionaries for sparse representation," *IEEE Trans. Signal Process.*, vol. 54, no. 11, pp. 4311–4322, Nov. 2006.

[14] S. Chen, D. Donoho, and M. Saunders, "Atomic decomposition by basis pursuit," *SIAM J. Sci. Comp.*, vol. 20, no. 1, pp. 33–61, 1998.

[15] S. Mallat and Z. Zhang, "Matching pursuits with time-frequency dictionaries," *IEEE Trans. Signal Process.*, vol. 41, no. 12, pp. 3397–3415, Dec. 1993.

[16] Y. C. Pati, R. Rezaiifar, and P. S. Krishnaprasad, "Orthogonal matching pursuit: Recursive function approximation with applications to wavelet decomposition," in *Proc. 1993 Conf. Rec. 27th Asilomar Conf. Signals, Systems Computers*, Pacific Grove, CA, 1993, pp. 40–44.

[17] J. A. Tropp, "Greed is good: Algorithmic results for sparse approximation," *IEEE Trans. Info. Theory*, vol. 50, no. 10, pp. 2231–2242, Oct. 2004.

[18] P. J. Huber and E. M. Ronchetti, *Robust Statistics*, 2nd ed. Hoboken, NJ: Wiley, 2009.

[19] S. Biswas, G. Aggarwal, and R. Chellappa, "Robust estimation of albedo for illumination-invariant matching and shape recovery," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 29, no. 3, pp. 884–899, Mar. 2009.

[20] V. Blanz and T. Vetter, "Face recognition based on fitting a 3D morphable model," *IEEE Trans. Pattern Ana. Machine Intell.*, vol. 25, no. 9, pp. 1063–1074, Sep. 2003.

[21] K. Lee, J. Ho, and D. J. Kriegman, "Acquiring linear subspaces for face recognition under variable lighting," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 27, no. 5, pp. 684–698, May 2005.

[22] S. Biswas and R. Chellappa, "Pose-robust albedo estimation from a single image," in *Proc. IEEE Conf. Computer Vision Pattern Recognition*, 2010.

[23] Q. Zhang and B. Li, "Discriminative K-SVD for dictionary learning in face recognition," in *Proc. IEEE Conf. Computer Vision Pattern Recognition*, San Francisco, CA, Jun. 2010, pp. 2691–2698.

[24] A. S. Georghiades, P. N. Belhumeur, and D. J. Kriegman, "From few to many: Ilumination cone models for face recognition under variable lighting and pose," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 23, no. 6, pp. 643–660, Jun. 2001.

[25] A. Martinez and R. Benavente, AR Face Database CVC Tech. Rep. 24, 1998.

[26] T. Sim, S. Baker, and M. Bsat, "The CMU pose, illumination, and expression database," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 25, no. 12, pp. 1615–1618, Dec. 2003.

[27] J. Ni and R. Chellappa, "Evaluation of state-of-the-art algorithms for remote face recognition," in *IEEE Intl. Conf. Image Processing*, Hong Kong, Sep. 2011.

[28] P. Viola and M. Jones, "Rapid object detection using a boosted cascade of simple features," in *Proc. IEEE Conf. Computer Vision Pattern Recognition*, Hilton Head, SC, Jun. 2001, pp. 511–518.

[29] G. R. Bradski and V. Pisarevsky, "Intel's computer vision library: Applications in calibration, stereo segmentation, tracking, gesture, face and object recognition," in *Proc. IEEE Conf. Computer Vision Pattern Recognition*, 2000, pp. 796–797.

[30] P. J. Phillips, "Support vector machines applied to face recognition," in *Advances in Neural Information Processing Systems 11*. Cambridge, MA: MIT Press, 1999, pp. 803–809.

[31] M. Deriche, "A simple face recognition algorithm using eigeneyes and a class-dependent PCA implementation," *Int. J. Soft Computing*, vol. 3, pp. 438–442, 2008.

[32] S. K. Zhou, G. Aggarwal, and D. W. Jacobs, "Appearance characterization of linear lambertian objects, generalized photometric stereo, and illumination-invariant face recognition," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 29, no. 2, pp. 230–245, Feb. 2007.

[33] S. Romdhani, V. Blanz, and T. Vetter, "Face identification by fitting a 3d morphable model using linear shape and texture error functions," in *Proc. Eur. Conf. Computer Vision*, Copenhagen, Denmark, May 2002, pp. 3–19.

[34] L. Zhang and D. Samaras, "Face recognition from a single training image under arbitrary unknown lighting using spherical harmonics," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 28, no. 3, pp. 351–363, Mar. 2006.

[35] P. Sinha, B. Balas, Y. Ostrovsky, and R. Russell, "Face recognition by humans: Nineteen results all computer vision researchers should know about," *Proc. IEEE*, vol. 94, no. 11, pp. 1948–1962, Nov. 2006.

[36] B. Leibe and B. Schiele, "Analyzing appearance and contour based methods for object categorization," in *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, 2003.

[37] F. Rodriguez and G. Sapiro, Sparse representations for image classification: Learning discriminative and reconstructive non-parametric dictionaries, Univ. Minnesota, Dec. 2007, Tech. Rep.

[38] E. Kokiopoulou and P. Frossard, "Semantic coding by supervised dimensionality reduction," *IEEE Trans. Multimedia*, vol. 10, no. 5, pp. 806–818, Aug. 2008.

[39] T. Zhang, Y. Y. Tang, B. Fang, Z. Shang, and X. Liu, "Face recognition under varying illumination using gradientfaces," *IEEE Trans. Image Process.*, vol. 18, no. 11, pp. 2599–2606, Nov. 2009.

[40] T. Chen, W. Yin, X. S. Zhou, D. Comaniciu, and T. S. Huang, "Total variation models for variable lighting face recognition," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 28, no. 9, pp. 1519–1524, Sep. 2006.

[41] K. Etemand and R. Chellappa, "Separability-based multiscale basis selection and feature extraction for signal and image classification," *IEEE Trans. Image Process.*, vol. 7, no. 10, pp. 1453–1465, Oct. 1998.

[42] K. Huang and S. Aviyente, "Sparse representation for signal classification," *Proc. Neural Information Processing Syst.*, vol. 19, pp. 609–616, 2007.

[43] M. Ranzato, F. Haung, Y. Boureau, and Y. LeCun, "Unsupervised learning of invariant feature hierarchies with applications to object recognition," in *Proc. IEEE Conf. Computer Vision Pattern Recognition*, Minneapolis, MN, 2007, pp. 1–8.

[44] M. Ranzato, C. Poultney, S. Chopra, and Y. LeCun, "Efficient learning of sparse representations with an energy-based model," in *Advances Neural Information Processing Systems*, Vancouver, B.C., Canada, 2006.

[45] J. Mairal, F. Bach, J. Pnce, G. Sapiro, and A. Zisserman, "Discriminative learned dictionaries for local image analysis," in *Proc. Conf. Computer Vision Pattern Recognition*, Anchorage, AL, Jun. 2008.

[46] J. Mairal, M. Leordeanu, F. Bach, M. Herbert, and J. Ponce, "Discriminative sparse image models for class-specific edge detection and image interpretation," in *Proc. Eur. Conf. Computer Vision*, Marseille, France, Oct. 2008.

[47] J. Mairal, F. Bach, J. Ponce, and G. Sapiro, "Online dictionary learning for sparse coding," in *Int. Conf. Machine Learning*, Montreal, Canada, June 2009.

[48] J. Mairal, F. Bach, J. Ponce, G. Sapiro, and A. Zisserman, "Supervised dictionary learning," in *Advances in Neural Information Processing Systems*, Vancouver, B.C., Canada, Dec. 2008.

[49] M. Gamassi, M. Lazzaroni, M. Misino, V. Piuri, D. Sana, and F. Scotti, "Quality assessment of biometric systems: A comprehensive perspective based on accuracy and performance measurement," *IEEE Trans. Instrumentation Measurement*, vol. 54, no. 4, pp. 1489–1496, Apr. 2005.

Vishal M. Patel (M'01) received the B.S. degrees in electrical engineering and applied mathematics (with honors) and the M.S. degree in applied mathematics from North Carolina State University, Raleigh, NC, in 2004 and 2005, respectively, and the Ph.D. degree in electrical engineering from the University of Maryland, College Park, in 2010.

In 2010, he was an ORAU Postdoctoral Fellow with the U.S. Army Research Laboratory. Currently, he is a member of the research faculty at the University of Maryland Institute for Advanced Computer Studies (UMIACS). His broad research interests are in signal processing, computer vision and pattern analysis with applications to biometrics recognition and radar imaging.

Dr. Patel is a member of Eta Kappa Nu, Pi Mu Epsilon, and Phi Beta Kappa.

Tao Wu (S'11) received the B.E. degree and the M.S. degree in electronics engineering from Tsinghua University, Beijing, respectively, in 2005 and 2008. He is currently working toward the Ph.D. degree in the Department of Electrical and Computer Engineering, University of Maryland, College Park.

His research interests include face recognition, computer vision, machine learning, and image and video processing.

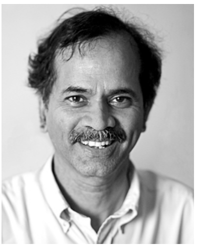
Soma Biswas (M'12) received the B.E. degree in electrical engineering from Jadavpur University, Kolkata, India, in 2001, the M.Tech. degree from the Indian Institute of Technology, Kanpur, in 2004, and the Ph.D. degree in electrical and computer engineering from the University of Maryland, College Park, in 2009.

She is currently working as a Research Assistant Professor at the University of Notre Dame, Notre Dame, IN. Her research interests are in signal, image, and video processing, computer vision, and pattern recognition.

P. Jonathon Phillips (M'96–SM'06–F'10) received the Ph.D. degree in operations research from Rutgers University, New Brunswick, NJ.

He is a leading Researcher in the fields of computer vision, face recognition, biometrics, and human identification. He is at the National Institute of Standards and Technology (NIST), where he directs challenge problems and evaluations in face recognition and biometrics. His efforts include the Multiple Biometrics Evaluation 2010, the Iris Challenge Evaluations (ICE), the Face Recognition Vendor Test (FRVT) 2006 and the Face Recognition Grand Challenge and FERET. From 2000 to 2004, he was assigned to the Defense Advanced Research Projects Agency (DARPA) as Program Manager for the Human Identification at a Distance program. He was Test Director for the FRVT 2002. His work has been reported in print media of record including the New York Times and the Economist. He has appeared on NPR's ScienceFriday.

For his work on the FRVT 2002, Dr. Phillips was awarded the Department of Commerce Gold Medal. From 2004 to 2008 he was an Associate Editor for the IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE and Guest Editor of an issue of the PROCEEDINGS OF THE IEEE on biometrics. In an Essential Science Indicators analysis of face recognition publication over the past decade, his work ranks at #2 by total citations and #1 by citations per paper. He is a Fellow of the IAPR.



Rama Chellappa (S'78–M'79–SM'83–F'92) received the B.E. (Hons.) degree from University of Madras, India, in 1975 and the M.E. (distinction) degree from Indian Institute of Science, Bangalore, in 1977. He received M.S.E.E. and Ph.D. degrees in electrical engineering from Purdue University, West Lafayette, IN, in 1978 and 1981, respectively.

Since 1991, he has been a Professor of electrical engineering and an affiliate Professor of computer science at the University of Maryland, College Park. He is also affiliated with the Center for Automation Research (Director) and the Institute for Advanced Computer Studies (Permanent Member). In 2005, he was named a Minta Martin Professor of Engineering. Prior to joining the University of Maryland, he was an Assistant (1981 to 1986) and Associate Professor (1986 to 1991) and Director of the Signal and Image Processing Institute (1988 to 1990) at the University of Southern California, Los Angeles. Over the last 30 years, he has published numerous book chapters, peer-reviewed journal and conference papers. He has co-authored and co-edited books on MRFs face and gait recognition and collected works on image processing and analysis. His current research interests are face recognition, clustering and video summarization, 3-D modeling from video, image and video-based recognition of objects, events and activities, dictionary-based inference, compressive sensing, and hyper spectral processing. He holds two patents.

Prof. Chellappa has received several awards, including an NSF Presidential Young Investigator Award, four IBM Faculty Development Awards, an Excellence in Teaching Award from the School of Engineering at USC, and two paper awards from the International Association of Pattern Recognition. He received the Society, Technical Achievement and Meritorious Service Awards from the IEEE Signal Processing Society. He also received the Technical Achievement and Meritorious Service Awards from the IEEE Computer Society. At University of Maryland, he was elected as a Distinguished Faculty Research Fellow, as a Distinguished Scholar-Teacher, and received an Outstanding Innovator Award from the Office of Technology Commercialization, and an Outstanding GEMSTONE Mentor Award. He received the Outstanding Faculty Research Award and the Poole and Kent Teaching Award for the Senior Faculty from the College of Engineering. In 2010, he was recognized as an Outstanding ECE by Purdue University. He is a Fellow of the International Association for Pattern Recognition, the Optical Society of America and the American Association for Advancement of Science. He served as the Associate Editor of four IEEE Transactions, as a Co-Editor-in-Chief of Graphical Models and Image Processing and as the Editor-in-Chief of IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE. He has also served as a Co-Guest Editor for five special issues published by IEEE Transactions, the PROCEEDINGS OF THE IEEE and the *International Journal of Computer Vision*. He served as a member of the IEEE Signal Processing Society Board of Governors and as its Vice President of Awards and Membership. He has served as a General and Technical Program Chair for several IEEE international and national conferences and workshops. He is a Golden Core Member of the IEEE Computer Society and served a two-year term as a Distinguished Lecturer of the IEEE Signal Processing Society. Recently, he completed a two-year term as the President of the IEEE Biometrics Council.