Research Article

Face Recognition Based on Optimized Projections for Distributed Intelligent Monitoring Systems

Aihua Yu,1 Huang Bai,1 Binbin Sun,1 Gang Li,1 and Beiping Hou2

1Zhejiang Provincial Key Laboratory for Signal Processing, College of Information Engineering, Zhejiang University of Technology, Hangzhou, Zhejiang 310023, China
2School of Automation and Electrical Engineering, Zhejiang University of Science and Technology, Hangzhou, Zhejiang 310023, China

Correspondence should be addressed to Aihua Yu; yuaihua_seu@163.com

Received 17 September 2015; Revised 29 January 2016; Accepted 18 February 2016

Academic Editor: Federico Berrero

Copyright © 2016 Aihua Yu et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Compressive sensing (CS), as a new theory of signal processing, has found many applications. This paper deals with a CS-based face recognition system design. A novel framework, called projection matrix optimization- (PMO-) based compressive classification, is proposed for distributed intelligent monitoring systems. Unlike the sparse preserving projection (SPP) approach, the projection matrix is designed such that the coherence between different classes of faces is reduced and hence a higher recognition rate is expected. The optimal projection matrix problem is formulated as identifying a matrix that minimizes the Frobenius norm of the difference between a given target Gram and that of the equivalent dictionary. A class of analytical solutions is derived. With the PMO-based CS system, two frameworks are proposed for compressive face recognition. Experiments are carried out with five popularly utilized face databases (i.e., ORL, Yale, Yale Extend, CMU PIE, and AR) and simulation results show that the proposed approaches outperform those existing compressive ones in terms of the recognition rate and reconstruction error.

1. Introduction

Face recognition (FR) has played a very important role in multimedia based applications. In spite of many years’ research, it remains an interesting and challenging research area [1]. Figure 1 depicts the conventional face recognition process. As it involves storing and transmitting high dimensional images, image compression techniques such as JPEG and JPEG2000 are used to alleviate the problem [2, 3]. At the receiver end, users have to decompress (reconstruct) images and extract the image features for classification. Such a procedure usually requires a lot of computations and hence makes the systems expensive. It can be much simplified if the images can be acquired using compressive sensing, which outputs features of images extracted directly, and the classification and reconstruction are done with the extracted features. See Figure 2.

With the development of Internet of Things and information technology, the demand for the distributed intelligent image monitor systems, as shown in Figure 3, increases greatly, which need an intelligent Internet end with the capacity of sensing and classifying. The cameras are connected by wireless Internet. The main problems with such a system are the transmission and storage of the images and the cost, including the infrastructure and the communications. To implement this architecture, technically speaking, there are several factors to consider: data storage space, RAM space, the computation time, and the transmission bandwidth. The key to solve these issues is to develop a data compression algorithm which can integrate the signal reconstruction and the classification with an accepted performance.

Dimension reduction plays an extensively important role in high dimensional data analysis and studies. In recent years, many dimensionality reduction methods are successfully applied in pattern recognition [4, 5]. The principal component analysis (PCA) intends to represent faces by projecting the facial images to the directions of maximal covariance in the facial image data. One of the advantages of PCA is to reduce dimensionality of data but such an approach ignores the relationship between data in high dimensions.
discrimination analysis (LDA) explicitly attempts to model the difference between the classes of data. The Fisherface method combines PCA and the Fisher criterion to extract the information that discriminates the differences between the classes of a sample set. The projection matrix is chosen to maximize the ratio between the determinant of the between-class scatter matrix of the projected samples and that of the within-class scatter matrix. Nevertheless, Martinez et al. demonstrated that when the training data set is small, the eigenface method outperforms the Fisherface method. Some novel algorithms attempt to reduce the data dimensionality, while keeping the intrinsic characteristics. The local preserve project (LPP) aims to find embedding which can preserve local information and obtain a face subspace that best detects the essential face manifold structure. Because LPP has an isometric property (RIP). See [6, 7].

The recently developed compressed sensing (CS) is a signal processing technique that can acquire a signal efficiently and reconstruct it by finding the solution to an underdetermined linear system [6, 7]. Its essence is to achieve analog signal discretization with sampling-compression integration, namely, the analog-to-discrete CS. The basic principle of such a CS framework is similar to that of discrete-to-discrete CS [8, 9], which can be explained below. In the standard CS framework, it is assumed that the high dimension signals \( x \in \mathbb{R}^{N \times 1} \) can be represented as a linear combination of \( L \) vectors \( \{\psi_k\} \):

\[
x = \sum_{k=1}^{L} s_k \psi_k = \Psi s,
\]

where \( \Psi \in \mathbb{R}^{N \times L} \) is known as the dictionary (matrix), while \( s \) is \( \|s\|_0 \)-sparse vector with \( \|s\|_0 \) denoting the number of nonzero elements of \( s \) and corresponding \( x \) is said to be \( \|s\|_0 \)-sparse in the dictionary \( \Psi \). The basic mathematical problem of CS is to study how to reconstruct the original high dimensional signal \( x \) from its low dimensional projection \( y \) which is mathematically of the form

\[
y = \Phi x,
\]

where \( \Phi \in \mathbb{R}^{M \times N} \) is called a projection matrix (it is also called measurement or sensing matrix, which will be used alternatively in this paper) with \( M < N \). Signal reconstruction means to find \( x \) from (2) with \( y \) and the pair \( (\Phi, \Psi) \) given. There are two conditions under which recovery is possible. The first one is sparsity which requires the signals \( x \) to be sparse in some \( \Psi \). The signal reconstruction problem is given as

\[
s \doteq \arg \min \limits_{s} \|s\|_0 \text{ s.t. } y = A\tilde{s},
\]

where \( A \doteq \Phi \Psi \) is called equivalent dictionary. The solution to (3) is unique for sparse signals if \( A \) satisfies the restricted isometric property (RIP). See [6, 7].

The second condition is related to the mutual coherence [10, 11] of the equivalent dictionary \( A \), which is defined below:

\[
\mu(A) \doteq \max_{1 \leq i \neq j \leq L} \frac{|A(:,i)^T A(:,j)|}{\|A(:,i)\|_2 \|A(:,j)\|_2},
\]

where \( ^T \) denotes the transpose operator. \( \mu(A) \) represents the worst-case coherence between any two atoms of \( A \). As shown in [10], the \( K \)-sparse signal can be exactly recovered from the measurement as long as

\[
\mu(A) < \frac{1}{2K-1}.
\]

As seen, the smaller the value of \( \mu(A) \) is, the bigger the value of \( K \) is allowed. The latter implies a wider range of signals that can be recovered exactly using such a CS system. So, minimizing \( \mu(A) \) is of importance. Simulations showed that the signal reconstruction accuracy is more related to the number of columns that are strongly correlated compared with \( \mu(A) \). Noting this fact, Eldad in [11] proposed to minimize averaged mutual coherence with respect to \( \Phi \) for a given dictionary \( \Psi \). Since then, many different approaches have appeared. Roughly speaking, these approaches are all based on the same framework: to design the sensing matrix \( \Phi \) such that the Gram matrix of \( A \), defined as \( A^T A \), is as close as to a target Gram matrix \( G_t \) as possible in the sense of

\[
\min_{\Phi, G_t \in \delta_t} \| G_t - A^T A \|_F^2,
\]

where \( \| \cdot \|_F \) denotes the Frobenius norm, \( \delta_t \) is a class of Gram matrices possessing certain properties, and the dictionary is...
assumed to be given. See [12–15]. As the diagonal elements of $G_t$ are all assumed to be one, the Frobenius norm-based error above actually represents the averaged coherence except for a constant factor $1/L(L-1)$ if all the diagonal entries of $A^T A$ are equal to one. The properties of the CS systems designed using (6) are strongly related to the choice of the target Gram $G_t$. A column normalized matrix $A \in \mathbb{R}^{M \times L}$ is said to be equiangular tight frame (ETF) if $|A(:, i)^T A(:, j)| = \sqrt{(L - M) / M (L - 1)}$, $\forall i \neq j$, where $\sqrt{(L - M) / M (L - 1)}$ is the smallest mutual coherence that an $M \times L$ matrix possibly has. Such a problem has been investigated for $G_t$ being the identity matrix and some analytical solutions are available [13]. In the ETF-based approach, the target Gram is taken as the one of a relaxed ETF matrix and the obtained sensing matrix shows an improved performance [13, 15]. Developing algorithms to solve (6) for arbitrary $G_t$ is still an interesting topic.

Sparse representation (SR), an important prerequisite for CS theory, has been applied in face classification [16]. Such an approach, usually referred to as SRC standing for sparse representation classification, can yield a higher recognition rate with varying illumination and expression. With SRC, the recognition is converted to the problem of classification among multiple linear regression models. The desired representation is sparse since the test sample should only be represented in terms of training samples belonging to the same class. The sparse representation can be computed with $l_1$ minimization [6]. However, this algorithm is very time-consuming for large scale databases with images of high resolution. The recognition rate degrades with the feature dimension reduction in general [16]. Sparse preserving projection (SPP), proposed in [17], builds the graphs based on the sparse reconstruction of data. Such a technique tries to preserve the sparsity without considering the coherence issue, which would affect the face recognition rate.

Sparse representation works well in applications where the original signal $x$ needs to be reconstructed as accurately as possible, such as denoising, image inpainting, and coding. However, sometimes we just need to discriminate the signal from its representation rather than reconstructing it. The difference between reconstruction and discrimination has been widely investigated in the literature. It is known that typical reconstructive methods, such as PCA and independent component analysis (ICA), aim at obtaining a representation that enables sufficient reconstruction, and thus they are able to deal with signal corruption, that is, noise, missing data, and outliers. On the other hand, discriminative methods, such as LDA, generate a signal representation that maximizes the separation of distributions of signals from different classes. While both classes of methods have broad applications in classification, the discriminative methods, as expected, have often outperformed the reconstructive methods for the classification task.

The main objective of this paper is to develop an algorithm for face classification and reconstruction, which is intended to be used in distributed monitoring systems and to be implemented in a low-cost microprocessor (e.g., ARM Cortex-M3 and ARM Cortex-M4) platform. The algorithm to be proposed in this paper is projection matrix optimization-(PMO-) based compressive classification, which tries to design the projection matrix $\Phi$ in such a way that the within-class coherence is enhanced while between-class one is reduced. The receiving end uses the sparse representation coefficients in the equivalent dictionary for image reconstruction and classification. Precisely speaking, the main contributions in this paper are given as follows:

(i) A new distributed monitoring system oriented face recognition framework is proposed based on compressed sensing. Instead of the high dimension original images, their lower dimension counterparts obtained using projection are transmitted and used for reconstruction and recognition. A new target Gram, denoted by $G_t$, is proposed for designing the projection matrix in order to improve the discrimination between classes.

(ii) The optimal projection matrix design problem is formulated in terms of identifying those sensing matrices that minimize the difference between the Gram of the equivalent dictionary and the proposed target Gram $G_t$ in Frobenius norm sense. A class of analytical solutions is derived for the proposed problem which is a generalization of that in [14].

(iii) With the PMO-based CS system, two frameworks are proposed for face recognition. Experiments are carried out and the results confirm that the proposed

**Figure 3**: Block-diagram of a distributed intelligent monitoring system.
approaches can effectively improve the system performance in terms of face classification and reconstruction.

The paper is outlined as follows. Section 2 is devoted to providing some existing works on compressive classification and recognition, which are closely related to ours. Our main contribution is given in Section 3, in which the PMO-based CS systems, two FR frameworks are proposed for distributed intelligent monitoring systems. Experiments are carried out in Section 4 to examine the performance of the proposed approaches and to compare them with some of the prevailing ones. To end this paper, some concluding remarks are given in Section 5.

2. Related Works and Problem Formulation

This section intends to review the sparse representation classification and some projection-based SRC methods, which have been considered as successful applications of signal sparse representation and compressive sensing theory to face recognition and are closely related to our present work to be developed in this paper.

2.1. Sparse Representation Classification. In [18], the sparse representation is shown to achieve state-of-the-art performance in image denoising. Such a technique was also used as an inpainting method in [19] for recovering missing pixels in images. It was extended to face recognition in [16].

Suppose we have P class samples \(\{C_p\}\); each class \(C_p\) has a set of samples with the same size, from which we randomly select \(Q\) samples for training purpose. Each sample forms a vector of dimension \(N \times 1\) and scaled to unit in \(L_2\)-norm, yielding an atom of the dictionary \(\Psi_p\):

\[
\Psi_p \triangleq [\Psi_1, \Psi_2, \ldots, \Psi_p, \ldots, \Psi_P] = [\psi_k],
\]

where \(\Psi_p \in \mathbb{R}^{N \times Q}\) is the dictionary for the \(p\)th class.

In the standard SR framework, it is assumed that the original face image signals \(x \in \mathbb{R}^{N \times 1}\) can be represented as a linear combination of \(PQ\) vectors \(\{\psi_k\}\):

\[
x = \sum_{k=1}^{L} s_k \psi_k \triangleq \Psi s,
\]

where \(L = PQ\) and \(s\) is sparse.

The procedure for classifying given \(L_2\)-normalized \(x\) contains the following steps [16].

\(\text{Alg}_{\text{SRC}}\)

**Step 1.** Normalize \(x\) in \(L_2\):

\[
x \triangleq \frac{x}{\|x\|_2}.
\]

**Step 2.** Find the sparse vector \(s\) with

\[
s = \arg \min_{\|s\|_1}
\]

\[
s.t. \quad x = \Psi s.
\]

Such a problem can be solved efficiently using a linear programming technique as the constraint is also linear.

**Step 3** (compute the residual energy). Denote by \(s_p \in \mathbb{R}^{Q 	imes 1}\) the subvector of \(s\) in \(\mathbb{R}^{L 	imes 1}\) corresponding to the \(p\)th dictionary \(\Psi_p\). Calculate

\[
\sigma_p^2 \triangleq \|x - \Psi_p s_p\|^2, \quad \forall p = 1, 2, \ldots, P.
\]

**Step 4** (classification). \(x\) belongs to the \(p\)th class, where \(\hat{p}\) is determined with

\[
\hat{p} \triangleq \arg \min_{p \in \{1, \ldots, P\}} \sigma_p^2.
\]

The input-output relationship of this algorithm is simply denoted by

\[
\hat{p} = \text{Alg}_{\text{SRC}}(\Psi, x),
\]

where the dictionary \(\Psi\) is \(L_2\)-normalized.

It should be pointed out that the SRC described above works on high dimensional signals \(x\). See (10). In the remainder of this section, we will present some methods that carry out the SRC in lower dimensionality domain by working with the projection \(y\) of original \(x\) via (2); that is, \(y = \Phi x\). Such a class of classification techniques is referred to as compressive SRC. Therefore, a method in this class differs from another by the ways in which the projection matrix \(\Phi\) is designed.

2.2. PCA-Based SRC. The PCA provides a dimensionality reduction technique to represent the high dimensional signal \(x\) with a lower dimensional one \(y\) obtained using (2). The projection matrix \(\Phi\) is determined as follows.

Let \(\bar{x} \triangleq (1/L) \sum_{k=1}^{L} x_k\) be the mean of \(\{x_k\}\). Denote \(e_k \triangleq x_k - \bar{x}\) and

\[
R \triangleq \sum_{k=1}^{L} e_k e_k^T.
\]

By singular value decomposition (SVD), \(R\) can be rewritten as

\[
R = U \Sigma U^T,
\]

where \(\Sigma = \text{diag}(\sigma_1, \ldots, \sigma_k, \ldots, \sigma_L)\) with the principal components satisfying \(\sigma_k \geq \sigma_{k+1}, \forall k\).

Under the PCA, the optimal projection matrix \(\Phi \in \mathbb{R}^{M \times N}\) with \(M \leq N\) is given by the first \(M\) columns of the orthonormal matrix \(U\):

\[
\Phi = U(:, 1 : M).
\]

The original signal \(x\) is projected into \(y = \Phi x\).
The PCA-based SRC approach, denoted by $\text{Alg}_{\text{PCA}}$, is a classification based on $y$ rather than $x$.

$\text{Alg}_{\text{PCA}}$

Step 1. With obtained $\Phi$ above, compute $y = \Phi x$ and $Y = \Phi \Psi$.

Step 2. Normalize the equivalent dictionary:

$$A \triangleq \text{Ydiag}(d_1, \ldots, d_k, \ldots, d_L),$$

where $d_k \triangleq 1/\|\text{Y}(k)\|_2$ with the denominator being the $l_2$-norm of the $k$th column of $Y$ for all $k \in \{1, 2, \ldots, L\}$.

Step 3. Determine the class that $x$ belongs to using

$$\hat{\rho} = \text{Alg}_{\text{SRC}}(A, y).$$

Such a procedure was proposed in [16].

2.3. SPP-Based SRC. The SPP proposed in [17] aims at carrying out the classification/recognition in a lower dimensional space. Precisely speaking, it works with signals $y$ which are obtained using (2), where the projection matrix $\Phi$ belongs to $M \times N$ with $M < N$.

In such a framework, the $k$th sample $x_k$ in the dictionary is featured by its sparse vector $\tilde{s}_k$ obtained with

$$\tilde{s}_k \triangleq \arg \min_{s \in \delta_k} \|s\|_1$$

s.t. $x_k = \Psi s_k$, $1 \leq 1^T s_k$,

where $1 \in \mathbb{R}^{L \times 1}$ is the column vector whose entries are all equal to 1 and $\delta_k$ is the vector space of dimension $L$, in which each of the vectors has its $k$th element equal to zero. Denote

$$E \triangleq \Psi - \Psi \tilde{s},$$

where $\tilde{s} \triangleq [\tilde{s}_1 \cdots \tilde{s}_1 \cdots \tilde{s}_L]$ as the sparse representation error matrix in the sense specified by (19). The projection matrix is chosen such that each row vector $\phi$ of $\Phi$ will minimize the sparse representation errors [17]:

$$\phi \triangleq \arg \min_{\phi} \|\phi E\|_2^2$$

s.t. $\phi \Psi \Psi^T \phi^T = 1,$

where the constraint is for avoiding degenerate solutions. Using Lagrange multiplier approach, one can find the solution to $\phi$ (21) from the following generalized eigenvalue problem:

$$E E^T v = \lambda \Psi \Psi^T v,$$

and the $M$ row vectors of optimal $\Phi$ are given the transposed eigenvectors, that is, $\{v_{1:k}\}$, corresponding to the top $M$ eigenvalues of the above equation. Equation (22) can be solved using MATLAB command eig.m:

$$[V, \Lambda] = \text{eig}(EE^T, \Psi \Psi^T),$$

where $\Lambda = \text{diag}(\lambda_1, \ldots, \lambda_k, \ldots, \lambda_L)$ with $\lambda_k \geq \lambda_{k+1}$ assumed. Then, the optimal projection matrix $\Phi \in \mathbb{R}^{M \times L}$ for SPP is given by first $M$ row vectors of $V^T$:

$$\Phi = (V^T (1 : M, :)).$$

As noted in [17], to avoid singularity of this problem, PCA-based preprocessing is usually used (let $XX^T = [W_1 W_2][\Sigma 0] [W_1 W_2]^T$ be the SVD of $XX^T$; then, replace $X$ with $\tilde{X} = W_1^T X$).

As understood, the SPP-based SRC method, denoted by $\text{Alg}_{\text{SRC}}$, is exactly the same as $\text{Alg}_{\text{PCA}}$ except that the projection matrix $\Phi$ is given by (24).

2.4. Sparse Related Face Recognition. In [20], a class of structured sparsity-inducing norms was included into SRC framework which mainly concerns the misalignment, shadow, and occlusion scenarios. But the authors did not take dimensionality reduction into account which is crucial for the implementation of distributed systems. SRC assumes that the sparse representation residual follows Gaussian or Laplacian distribution, while robust sparse coding (RSC) considered in [21] models the sparse coding as a sparsity constrained robust regression problem and seeks the maximum likelihood estimation solution of the sparse coding. All these make it much more robust than SRC. However, the computational complexity of RSC is also much higher than SRC's which is not practical for the investigated distributed system.

2.5. Problem Formulation. In a distributed intelligent monitoring system, $N_1 \times N_2$ images $\{I_k\}$ acquired by the agents are usually of high dimension and have to be transmitted to the back-end server for recognition. In order to have an efficient transmission, it is desired to compress these high dimension signals $\{x_k\}$ with $x_k \in \mathbb{R}^{N \times 1}$, where $N \approx N_1 N_2$. One way to do so is through projection $y_k = \Phi x_k$, where the projection/sensing matrix $\Phi \in \mathbb{R}^{M \times N}$ with $M \ll N$ compresses $x_k$ into a much lower dimension signal $y_k$. Instead of $x_k$, $y_k$ is transmitted.

Once $y_k$ is received by the back-end server, the recognition of $y_k$ (corresponding to image $I_k$) is done with the equivalent dictionary $A$:

$$A = \Phi \Psi = [\Phi \Psi_1 \cdots \Phi \Psi_p \cdots \Phi \Psi_p],$$

where the dictionary $\Psi$, as formed in (7), is assumed to be given. The main problem to be investigated in the remainder of this paper is how to design the sensing matrix $\Phi$ such that the image $I_k$ can be recognized with the measurement $y_k$ and the equivalent dictionary $A$.

3. The Proposed CS System for FR Distributed Monitoring Systems

As mentioned before, the key to solve the problems encountered in distributed intelligent monitoring systems is to
compress the original (high dimensional) image signals and to carry out face recognition effectively with the low dimensional measurements.

This can be done using CS systems. In fact, both Alg_{SRC} and Alg_{SRC} from Section 2, can be viewed as two special CS systems in which the sensing matrix is designed to preserve principal components of the signal covariance matrix and the sparse representation of the samples in the dictionary $\Psi$, respectively.

As seen in the previous section, the optimal sensing matrix design problem is formulated by (6). Such a problem has been studied intensively for signal compression application, in which achieving high reconstruction accuracy is the ultimate goal of the design.

In the context of face recognition, what we are interested in the most is the recognition rate. Therefore, it is desired to design the projection matrix $\Phi$ to enhance the recognition rate. This is strongly related to how to choose the target Gram $G_t$. Another important issue is how to solve (6) with arbitrary $G_t$.

### 3.1. How to Choose the Target Gram?

Optimizing projection matrix for CS systems has been an important research topic in the context of face recognition. It was initiated by Elad’s work reported in 2007 [11]. Since then, many theoretical achievements have been obtained [12–15]. Roughly speaking, all these results are based on the following formulation:

\[
\min_{\Phi} \| G - G_t \|_F^2 \\
\text{s.t. } G = A^T A,
\]

where $\| \cdot \|_F$ is defined as the Frobenius norm, $G$ is the Gram matrix of the equivalent dictionary $A = \Phi \Psi$, and $G_t$ is a target Gram matrix.

For a given dictionary, the performance of the projection matrix depends on the choice of $G_t$. When a CS system is designed for signal compression, the reconstruction accuracy is the ultimate goal to achieve. In such a case, $G_t$ is chosen such that the resultant projection matrix makes the equivalent dictionary $A$ have small coherence between its columns and the system robust against noises, say sparse representation errors of signals. Recently, Cleju showed in [14] that the images can be well reconstructed from the low dimensional measurements projected with the sensing matrix $\Phi$ designed based on $G_t = \Psi^T \Psi$, the Gram of the dictionary.

**Comment 1.** It should be pointed out that, in $N_1 \times N_2$-image compression oriented CS systems, the images are divided into $n_1 \times n_2$ subimages, each of which forms a signal vector, say $x_k \in \mathbb{R}^{n_1 n_2}$ with $k \equiv n_1 n_2$ (this is very different from what has been used in the SRC-based face recognition approaches, in which the signals are of dimension $N_1 N_2 \times 1$), while the constructive dictionary $\Psi$ is of dimension $I \times L$ with $L \gg I$.

The K-SVD [18] has been considered as one of the most successful methods for designing such a dictionary. It was observed that the dictionaries obtained with different classes of images/samples are very similar. And our experiments showed that a direct application of such a CS system to face recognition based on reconstruction error does not yield a satisfactory performance in terms of recognition rate. Note that, for face recognition application, our ultimate objective is accurate recognition and to enhance recognition rate it would be better to design the sensing matrix $\Phi$ by improving the discrimination between the classes at the price of sacrificing reconstruction accuracy. This is actually one of the main motivations for our proposed approach in this paper, which will be developed further.

Let $\Psi = [\Psi_1 \cdots \Psi_p \cdots \Psi_p]$ with $\Psi_p \in \mathbb{R}^{N \times Q}$ be the overall dictionary defined before for face recognition and its Gram is

\[
G_p \equiv \Psi^T \Psi = \begin{bmatrix}
G_{11} & G_{12} & \cdots & G_{1p} \\
G_{21} & G_{22} & \cdots & G_{2p} \\
\vdots & \vdots & \ddots & \vdots \\
G_{p1} & G_{p2} & \cdots & G_{pp}
\end{bmatrix}
\]

with $G_{ij} \equiv \Psi_i^T \Psi_j$.

When choosing $\Phi \Psi$ such that $\| G_p - G \|_F^2$ is minimized, one emphasizes the reconstruction accuracy as the fundamental assumption of the SRC-based face recognition approach is that a sample (image), belonging to a class, is close to a linear combination of a few atoms of the dictionary of the same class. This, however, can not ensure that such a sample can not be well represented sparsely in a linear combination of other (sub)dictionaries. In other words, the reconstruction error itself may not be a proper measure for recognition. As mentioned before, recognition rate is our primary goal and hence in the CS system design, it is desired to choose the projection matrix $\Phi$ such that the coherence between the columns of $A_t$ and those of $A_j$, for all $i \neq j$ is reduced. By doing so, the discrimination between classes is improved. This can be achieved by choosing the sensing matrix such that the Gram of the corresponding equivalent dictionary is close to the following target Gram as much as possible:

\[
G_t = G_p + \Delta,
\]

where

\[
\Delta \equiv \begin{bmatrix}
\Delta_{11} & \Delta_{12} & \cdots & \Delta_{1p} \\
\Delta_{21} & \Delta_{22} & \cdots & \Delta_{2p} \\
\vdots & \vdots & \ddots & \vdots \\
\Delta_{p1} & \Delta_{p2} & \cdots & \Delta_{pp}
\end{bmatrix}
\]

with $\Delta_{ij}$ defined below:

\[
\Delta_{ij} (m, n) = \begin{cases}
-w \ast \eta, & i \neq j \\
\eta, & i = j, \ m \neq n \\
0, & i = j, \ m = n,
\end{cases}
\]
where both $\eta > 0$ and $\omega > 0$ are used to adjust the coherence between different classes and that between the columns of the same class, called correction parameter. These parameters should be chosen such that the off-diagonal elements of $G_i$ are all within $[0, 1]$. Such a target Gram $G_i$ is called discriminative Gram.

**Comment 2.** (i) In [12], sensing matrix optimization for block-sparse decoding was investigated. It is observed that our problem is the simplest block-sparse case, where the blocks are prefixed according to the classes of face images. $G_i$ is taken as the identity matrix in [12] and in our problem is assumed to be symmetric. Both encounter how to choose the weighting factors, which is application dependent and purely empirical. We will present some experimental results in the next section.

(ii) It should be pointed out that though, as to be seen, the recognition/classification is done based on the errors $\|y - A_i s\|_F^2$, which has a physical meaning close to the reconstruction error, these errors have taken the discrimination between classes into account via all $\{A_i\}$ or the sensing matrix which is designed based on the discriminative Gram $G_i$ defined in (28). So, such a measure can be considered as a mixture of reconstruction error and discriminative error. It is worth pointing that, in [22], a dictionary learning-based classification scheme was investigated where the dictionary was trained by certain class of signals as the classifier, while, in this paper, we consider the classification task in compressive domain. The sensing matrix is designed according to the given dictionary and hence the equivalent dictionary (corresponding to the compressive domain) possesses good property so that the sparse coding can be implemented accurately.

In the next subsection, we will discuss how to solve the proposed optimal sensing matrix design problem.

### 3.2. An Algorithm for Optimizing Projection Matrix

With the target Gram $G_i$ defined in (28), one can now consider the optimal projection matrix design, which is formulated as follows:

$$
\Phi = \arg\min_\Phi \| G - G_i \|_F^2 \\
\text{s.t. } G = \Psi^T \Phi^T \Phi \Psi.
$$

(31)

Now, let us consider how to solve the above problem. First of all, assume that $\Psi$ has the following SVD:

$$
\Psi = U_\Psi \left[ \begin{array}{cc}
\Sigma_\Psi & 0 \\
0 & 0
\end{array} \right] V_{\Psi}^T,
$$

(32)

where $\Sigma_\Psi \in \mathbb{R}^{N \times N} > 0$. Then, we get

$$
G = V_\Psi \left[ \begin{array}{cc}
\Sigma_\Psi & 0 \\
0 & 0
\end{array} \right] W \left[ \begin{array}{cc}
\Sigma_\Psi & 0 \\
0 & 0
\end{array} \right] V_{\Psi}^T,
$$

(33)

where $W \triangleq U_{\Psi}^T \Phi^T \Phi U_\Psi$. Let

$$
W = \begin{bmatrix}
W_{11} & W_{12} \\
W_{21} & W_{22}
\end{bmatrix}
$$

(34)

s.t. $W_{11} \in \mathbb{R}^{N \times N}$.

Then,

$$
\| G - G_i \|_F^2 = \| V_\Psi \left[ \begin{array}{cc}
\Sigma_\Psi & 0 \\
0 & 0
\end{array} \right] W \left[ \begin{array}{cc}
\Sigma_\Psi & 0 \\
0 & 0
\end{array} \right] V_{\Psi}^T - G_i \|_F^2
$$

$$
\triangleq \| \Sigma_\Psi W_{11} \Sigma_\Psi - G_i \|_F^2,
$$

where

$$
G_i \triangleq V_{\Psi}^T G_i V_{\Psi} = \begin{bmatrix}
G_{11} & G_{12} \\
G_{21} & G_{22}
\end{bmatrix}
$$

(36)

s.t. $G_{11} \in \mathbb{R}^{N \times N}$.

Finally, the cost function can be rewritten as

$$
\| G - G_i \|_F^2 = \| \Sigma_\Psi W_{11} \Sigma_\Psi - G_i \|_F^2 + \| G_{21} \|_F^2 + \| G_{11} \|_F^2.
$$

(37)

We can see that the two right-hand sides have nothing to do with the projection matrix. Define

$$
\bar{W}_{11} \triangleq \Sigma_\Psi W_{11} \Sigma_\Psi.
$$

(38)

Equation (31) is equivalent to

$$
\Phi = \arg\min_\Phi \| \bar{W}_{11} - G_{11} \|_F^2.
$$

(39)

Let $\bar{W}_{11} = V_{W} A_{W} V_{W}^T$ and $G_{11} = V_{A} A_{1} V_{A}^T$ be an SVD and eigen decomposition of $\bar{W}_{11}$ and $G_{11}$, respectively. Furthermore, assume that both $\{A_{W}(k,k)\}$ and $\{A_{i}(k,k)\}$ are in descending order. It then follows from [23] (see Corollary 7.4.9.3 in page 468) that

$$
\| \bar{W}_{11} - G_{11} \|_F^2 \geq \| A_{W} - A_{i} \|_F^2 \triangleq \varrho
$$

(40)

and the lower bound (equality) is achieved if and only if

$$
V_{W} = V_{i}.
$$

(41)

Noting that rank($\bar{W}_{11}$) $\leq M$, one can see that the lower bound $\varrho$ can be minimized with

$$
A_{i}(k,k) = \begin{cases}
A_{i}(k,k), & \forall k \leq M \text{ that } A_{i}(k,k) \geq 0 \\
0, & \forall k \leq M \text{ that } A_{i}(k,k) < 0.
\end{cases}
$$

(42)

With (41) and (42), one has optimal $\bar{W}_{11}$ and hence $W_{11}$:

$$
\bar{W}_{11} = V_{i} A_{W} V_{i}^T \implies W_{11} = \Sigma_\Psi^{-1} \bar{W}_{11} \Sigma_\Psi^{-1}.
$$

(43)
Let the following be an SVD of $W_{11}$:

$$W_{11} = V_{11} \begin{bmatrix} \Sigma_{11}^2 & 0 \\ 0 & 0 \end{bmatrix} V_{11}^T.$$  

(44)

And set $W_{12} = 0$, $W_{21} = 0$, and $W_{22} = 0$; then

$$W = \begin{bmatrix} V_{11} & 0 \\ 0 & V_{22} \end{bmatrix} \begin{bmatrix} \Sigma_{11}^2 & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} V_{11} & 0 \\ 0 & V_{22} \end{bmatrix}^T,$$  

(45)

where $V_{22}$ is any orthonormal matrix with dimensions $(N - \bar{N})$.

Since $\Phi^T \Phi = U_p W U_p^T$, a class of solutions to (31) is obtained as

$$\tilde{\Phi} = U \begin{bmatrix} \Sigma_{11} & 0 \\ 0 & V_{22} \end{bmatrix} \begin{bmatrix} V_{11} & 0 \\ 0 & V_{22} \end{bmatrix}^T U_p^T,$$  

(46)

where $\Sigma_{11} \in \mathbb{R}^{M \times M}$ and $U$ is any orthogonal matrix with dimension of $M$.

Seen from the above discussion, the optimization of projection matrix $\Phi$ is only related to the dictionary $\Psi$ and the correction constants $w$ and $\eta$, and, therefore, for given $\Psi$, $w$, and $\eta$, we can get $\Phi$, which is better than traditional feature selecting process. And (46) is an analytical solution. In addition, there are two degrees of freedom $U$ and $V_{22}$ in the results, which provides the possibility to further improve system performance.

Comment 3. Our proposed optimal sensing design problem shares the same form as that in [14]. It should be pointed out that, in [14], $G_s = G_p$ and the solutions are relatively easy to obtain, while our $G_s$ is a generalization of $G_p$ and the solutions are actually applicable to any symmetric $G_s$. In terms of applications, ours is for face recognition, in which the degrees of freedom, brought by the parameters $\eta$ and $w$, make our approach outperform that in [14].

3.3. The Proposed CS-Based Frameworks for FR. With the obtained optimal sensing matrix $\Phi$, any given high dimension face image $x$ can be projected to a vector of much lower dimension, which will be transmitted to the server of the distributed monitoring system for classification/recognition. Two PMO-based classification methods are proposed here. The first one, denoted by $Alg_{PMO}^{SRC}$, has exactly the same structure as $Alg_{SPCA}^{SRC}$ and $Alg_{SPP}^{SRC}$, except that the projection matrix $\Phi$ is given by (46).

The second one, denoted by $Alg_{PMO}^{P}^{j}$, is outlined as follows.

$Alg_{PMO}^{P}^{j}$

Step 1. At a subsystem that captures a face image $x$, compress it using

$$y = \Phi x$$  

(47)

and then encode and transmit $y$ to the server of the distributed monitoring system.

Step 2. At the server that receives $y$, compute

$$\sigma_p^2 \triangleq \min_{\hat{y}} \| y - A_p \hat{y} \|_2^2, \quad p = 1, \ldots, P,$$  

(48)

where $A_p \in \mathbb{R}^{M \times Q}$ is the equivalent dictionary for the $p$th class: $A = \tilde{\Phi} \Psi = [A_1, \cdots, A_p, \cdots, A_P]$.

Step 3 (classification). $x$ belongs to the $\tilde{p}$th class, where $\tilde{p}$ is determined with

$$\tilde{p} \triangleq \arg \min_{p \in \{1, \ldots, P\}} \{ \sigma_p^2 \}.$$  

(49)

In the next section, we will examine the performance of the proposed algorithms and compare it with some existing ones.

4. Experiment Results

In this section, we examine the performance of various classification systems on five face databases: ORL [24], Yale [25], Yale Extend [26] (referred to as Yale-E), CMU PIE [27] (referred to as PIE), and AR [28]:

(i) ORL database contains 400 images of 40 individuals (each provides 10 images); that is, $P = 40$. Some images are captured at different time, varying the lighting, facial expressions, and facial details. In our test, we randomly select $Q = 5$ images from each individual to form the dictionary set, while the remaining ones are for testing.

(ii) Yale database contains 165 images of 15 individuals (i.e., $P = 15$; each provides 11 images) under various illumination conditions and facial expressions. A random subset with $Q = 5$ images per individual is taken to form the dictionary set, and the rest of the database is used for testing.

(iii) Yale-E database includes the Yale face database B and the extended Yale face database. A subset called Yale Extend face database is collected from these two databases, which contains 2414 face images of 38 subjects; that is, $P = 38$. In our test, we randomly select $Q = 20$ images from each individual to form the dictionary set, while keeping the remaining one for testing.

(iv) PIE database is composed of 68 subjects (i.e., $P = 68$) with 41368 face images which are captured by 13 synchronized cameras and 21 flashes, under varying pose, illumination, and expression. We generate a data set from it by selecting 184 images for each individual and totally 12512 samples are used in our experiments. $Q = 20$ images are sampled to form the dictionary set, while the remaining images are considered for testing.

(v) AR database contains over 4000 color face images of 126 people (70 men and 56 women; i.e., $P = 126$) taken during two distinct photo sessions (separated by two weeks), with different facial expression, lighting conditions, and occlusions. We choose 50 men and 50 women to generate a data set of 100
persons (each provides 26 images). Random selection of $Q = 20$ images per individual is taken to form the dictionary set, and the rest of the database is for testing.

For each of the five databases, we take all $P$ classes (different persons) with each class containing $Q$ different samples (images), leading to $L = PQ$. All images are resized to resolution of $32 \times 32$ and normalized in scale; hence, each face sample is represented as a column vector $x \in \mathbb{R}^{N \times 1}$ with $N = 1024$. By doing so, the corresponding dictionary $\Psi \in \mathbb{R}^{N \times L}$ is then generated.

The systems consisted of different compression methods (PCA, SPP, RDM for random sampling matrix, and PMO) and various classifiers (SRC, RSC, Wlearning for [22], and $l_2$) are compared. For instance, $\text{Alg}_{\text{PCA}}^{\text{SRC}}$ denotes the system combining PCA and SRC with the subscript referring to compression method and the superscript the classifier.

4.1. The Effect of $Q$. Firstly, we briefly discuss the effect of the number of training samples with ORL database. Figure 4 shows the results of $\text{Alg}_{\text{PMO}}^{l_2}$ versus $Q$ with different $\eta$ values and $w = 2$, $M = 80$.

As can be seen, an increasing trend is obtained when $Q$ gets bigger in general with slight fluctuation. It is reasonable that more samples are used for representation, and higher accuracy is achieved. In the following, proper $Q$ values will be chosen for different databases as introduced in the beginning of this section.

4.2. The Choice of the Correction Parameter $\eta$. One highlight of the proposed PMO method is the correction parameter $\eta$. Now, we set up experiment for testing the effect of $\eta$. Fixing $M = 80$ and $w = 2$, the recognition rate versus the correction parameter $\eta$ is depicted in Figure 5.

It is clearly seen from the figure that the correction parameter $\eta$ does affect the performance of classification systems though the value should be chosen as database adapted.

4.3. Comparison of Different Classifiers. In this part, we turn to examine the performance of classification methods (i.e., SRC, RSC, $l_2$, and Wlearning) with PMO applied for compression. Figure 6 shows the recognition results with ORL database used.

The computation times of different methods are given in Table 1.

For this case, the proposed $l_2$ classifier achieves the best results almost in every $\eta$. The recognition rates of RSC are comparable somewhere, but its computation complexity is
Table 1: Computation time of different classifiers.

| Alg          | SRC  | RSC  | PMO | $l_2$PMO |
|--------------|------|------|-----|----------|
| Time (s)     | 4.39 | 29.80| 1.96| 1.87     |

Figure 7: Recognition rates versus $M$.

much higher than the others as can be concluded from Table 1, and this is also why RSC is not suitable for large distributed system.

4.4. Comparison of Different Compression Methods. The scenarios where various compression methods are combined with the SRC are investigated and the statistical results are summarized in Table 2.

The computation time with ORL database considered is given in Table 3.

The data in Table 2 demonstrates the superiority of the proposed PMO methods. The computation times of various systems differ with each other in the projection matrix design methods, that is, dimensionality reduction procedure. PMO costs more time when compared with PCA. But the design process is an offline job, and it does not put any additional pressure on the front agents.

4.5. Recognition Rate versus Compression Dimension. In this part, we examine the performance of different systems with varying compression rates on ORL database for $\eta = 0$, $\nu = 2$, and $Q = 5$. Four compression methods are conducted with SRC. In addition, PMO combined with the proposed $l_2$ classifier is also carried out for comparison. Figure 7 displays the recognition accuracies versus measurement dimension $M$.

As can be seen, higher recognition rates are obtained when bigger $M$ values are set. The results in Figure 7 are coincident with the statistics of Table 2 that PMO achieves the best recognition rate, which is better than the results of PCA and RDM, and SPP once again performs the worst.

4.6. Effect of Occlusion. One of the most challenging problems in FR techniques is the robustness to face occlusion. In this subsection, we test the performance of our proposed method with different occlusion scenarios. The AR database consists of occlusion face images. Taking these samples for test, Figure 8 shows the recognition rates versus measurement dimension.

This figure indicates that even for occlusion scenarios, as the measurement dimension increases, better recognition rate results can be achieved and SPP is the worst of all.

Figure 9 depicts the recognition rates versus the occlusion percentage with ORL database used, and $M = 80$, $\eta = 0.02$, and $\nu = 2$ for the proposed PMO.

The results of Figure 9 demonstrate that SPP and RDM are more sensitive to the occlusion, and PMO and PCA always perform better than the other two.

4.7. Image Reconstruction Experiment. The PMO algorithm is applied to distributed intelligent monitoring system, and image reconstruction can be performed on the server terminal. Denote the test sample $x \in \mathbb{R}^{N \times 1}$ and the reconstructed signal $\hat{x} \in \mathbb{R}^{N \times 1}$. The mean square error (MSE) is defined as

$$
\sigma_{MSE} \triangleq \frac{1}{N} \| \hat{x} - x \|^2_2.
$$

A popular used indicator to evaluate the image reconstruction accuracy is peak signal-to-noise ratio (PSNR) defined as

$$
\sigma_{PSNR} \triangleq 10 \times \log_{10} \left( \frac{(2^r - 1)^2}{\sigma_{MSE}} \right)
$$

with $r = 8$ bits/pixel. Figure 10 shows the evaluation of $\sigma_{PSNR}$ for varying measurement dimension $M$ with $\eta = 0.02$ and $\nu = 2$ for the proposed PMO.
Table 2: Comparison of different compression methods with $M = 80$, $\eta = 0.02$, and $\omega = 2$.

| Database | ORL    | Yale   | Yale-E  | PIE    | AR     |
|----------|--------|--------|---------|--------|--------|
| AlgSRC   | 92.20% | 85.56% | 91.63%  | 86.58% | 93.60% |
| AlgPCA   | 93.53% | 92.22% | 92.21%  | 89.71% | 95.80% |
| AlgSPP   | 46.75% | 94.44% | 44.36%  | 21.18% | 74.46% |
| AlgPMO   | 95.00% | 93.26% | 91.89%  | 90.68% | 98.24% |

Figure 9: Recognition rate versus occlusion percentage with ORL database tested.

Again, the results of PMO and PCA are similar and better than the others. Please note that as AlgSRC_SPP performs much worse than the others, we omit it in this figure. One example is given in Figure 11 to demonstrate the visual effect of the proposed system.

5. Conclusion

This work presents a novel CS-based face recognition scheme for distributed intelligent monitoring systems. A new compression strategy has been proposed based on the projection matrix optimization. The analytical solution set of the corresponding optimization problem has also been derived. With this new compression method, two frameworks have been presented for compressive face recognition. Experimental results on face recognition tasks demonstrate the superiority of the proposed approaches.

The monitor system always encounters the problem that we need to register new face and update the whole system. The FR oriented online CS system learning is an interesting direction of research for distributed intelligent monitor system design. In our approach, the dictionary is formed directly with the samples. It is expected that more efficient FR oriented CS systems can be achieved if both the dictionary and the sensing matrix are optimized alternatively or simultaneously.
Competing Interests
The authors declare that they have no competing interests.

Acknowledgments
This work was supported by NSFC Grants 61273195, 61304124, 61473262, and 61503339 and ZJNSF Grants LY13F010009 and LQ14F030008.

References
[1] W. Y. Zhao, R. Chellappa, P. J. Phillips, and A. Rosenfeld, “Face recognition: a literature survey,” ACM Computing Surveys, vol. 35, no. 4, pp. 399–458, 2003.
[2] G. K. Wallace, “The JPEG still picture compression standard,” IEEE Transactions on Consumer Electronics, vol. 38, no. 1, 1992.
[3] D. S. Taubman and M. W. Marcellin, JPEG2000 Image Compression Fundamentals, Standards and Practice, Kluwer Academic, New York, NY, USA, 2002.
[4] W. Chen, M. J. Er, and S. Wu, “PCA and LDA in DCT domain,” Pattern Recognition Letters, vol. 26, no. 15, pp. 2474–2482, 2005.
[5] M. Turk and A. Pentland, “Eigenfaces for recognition,” Journal of Cognitive Neuroscience, vol. 3, no. 1, pp. 72–86, 1991.
[6] D. L. Donoho, “Compressed sensing,” Transactions on Information Theory, vol. 52, no. 4, pp. 1289–1306, 2006.
[7] E. J. Candès and M. B. Wakin, “An introduction to compressive sampling,” IEEE Signal Processing Magazine, vol. 25, no. 2, pp. 21–30, 2008.
[8] Y. Liu, X. Zhu, L. Zhang, and S. H. Cho, “Distributed compressed video sensing in camera sensor networks,” International Journal of Distributed Sensor Networks, vol. 2012, Article ID 352167, 10 pages, 2012.
[9] W. Li, T. Jiang, and N. Wang, “Compressed sensing based on the characteristic correlation of ECG in hybrid wireless sensor network,” International Journal of Distributed Sensor Networks, vol. 2015, Article ID 325038, 8 pages, 2015.
[10] D. L. Donoho and M. Elad, “Optimally sparse representation in general (nonorthonormal) dictionaries via ℓ 1 minimization,” Proceedings of the National Academy of Sciences, vol. 100, no. 5, pp. 2197–2202, 2003.
[11] M. Elad, “Optimized projections for compressed sensing,” IEEE Transactions on Signal Processing, vol. 55, no. 12, pp. 5695–5702, 2007.
[12] L. Zelnik-Manor, K. Rosenblum, and Y. C. Eldar, “Sensing matrix optimization for block-sparse decoding,” IEEE Transactions on Signal Processing, vol. 59, no. 9, pp. 4300–4312, 2011.
[13] G. Li, Z. H. Zhu, D. H. Yang, L. P. Chang, and H. Bai, “On projection matrix optimization for compressive sensing systems,” IEEE Transactions on Signal Processing, vol. 61, no. 11, pp. 2887–2898, 2013.
[14] N. Cleju, “Optimized projections for compressed sensing via rank-constrained nearest correlation matrix,” Applied and Computational Harmonic Analysis, vol. 36, no. 3, pp. 495–507, 2014.
[15] H. Bai, G. Li, S. Li, Q. Li, Q. Jiang, and L. Chang, “Alternating optimization of sensing matrix and sparsifying dictionary for compressed sensing,” IEEE Transactions on Signal Processing, vol. 63, no. 6, pp. 1581–1594, 2015.
[16] J. Wright, A. Y. Yang, A. Ganesh, S. S. Sastry, and Y. Ma, “Robust face recognition via sparse representation,” IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 31, no. 2, pp. 210–227, 2009.
[17] L. Qiao, S. Chen, and X. Tan, “Sparsity preserving projections with applications to face recognition,” Pattern Recognition, vol. 43, no. 1, pp. 331–341, 2010.
[18] M. Aharon, M. Elad, and A. Bruckstein, “K-SVD: an algorithm for designing overcomplete dictionaries for sparse representation,” IEEE Transactions on Signal Processing, vol. 54, no. 11, pp. 4311–4322, 2006.
[19] T.-J. Chin and D. Suter, “Incremental kernel principal component analysis,” IEEE Transactions on Image Processing, vol. 16, no. 6, pp. 1662–1674, 2007.
[20] K. Jia, T.-H. Chan, and Y. Ma, “Robust and practical face recognition via structured sparsity,” in Computer Vision—ECCV 2012: 12th European Conference on Computer Vision, Florence, Italy, October 7–13, 2012, Proceedings, Part IV, vol. 7575 of Lecture Notes in Computer Science, pp. 331–344, Springer, Berlin, Germany, 2012.
[21] M. Yang, L. Zhang, J. Yang, and D. Zhang, “Robust sparse coding for face recognition,” in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR ’11), pp. 625–632, Providence, RI, USA, June 2011.
[22] J. Mairal, F. Bach, and J. Ponce, “Task-driven dictionary learning,” IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 34, no. 4, pp. 791–804, 2012.
[23] R. A. Horn and C. R. Johnson, Matrix Analysis, Cambridge University Press, Cambridge, UK, 2nd edition, 2012.
[24] F. S. Samaria and A. C. Harter, “Parameterisation of a stochastic model for human face identification,” in Proceedings of the 2nd IEEE Workshop on Applications of Computer Vision, pp. 138–142, December 1994.
[25] F. S. A. Georghiades, P. N. Belhumeur, and D. J. Kriegman, “From few to many: illumination cone models for face recognition under variable lighting and pose,” IEEE Transactions on
[26] K.-C. Lee, J. Ho, and D. J. Kriegman, "Acquiring linear subspaces for face recognition under variable lighting," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 27, no. 5, pp. 684–698, 2005.

[27] T. Sim, S. Baker, and M. Bsat, "The CMU pose, illumination, and expression (PIE) database," in *Proceedings of the 5th IEEE International Conference on Automatic Face Gesture Recognition (FGR ’02)*, pp. 53–58, Washington, DC, USA, May 2002.

[28] A. M. Martinez and R. Benavente, "The AR face database," CVC Technical Report 24, 1998.
