IET Image Processing

Research Article

IET Journals
The Institution of Engineering and Technology

Fast matching pursuit for sparse representation-based face recognition

ISSN 1751-9659
Received on 18th November 2017
Revised 2nd April 2018
Accepted on 20th April 2018
E-First on 13th June 2018
doi: 10.1049/iet-ipr.2017.1263
www.ietdl.org

Michael Melek[1], Ahmed Khattab[1] ✉, Mohamed F. Abu-Elyazeed[1]

[1]Electronics and Electrical Communications Engineering Department, Cairo University, 12613 Giza, Egypt

✉ *E-mail: akhattab@ieee.org*

Abstract: Even though face recognition is one of the most studied pattern recognition problems, most existing solutions still lack efficiency and high speed. Here, the authors present a new framework for face recognition which is efficient, fast, and robust against variations of illumination, expression, and pose. For feature extraction, the authors propose extracting Gabor features in order to be resilient to variations in illumination, facial expressions, and pose. In contrast to the related literature, the authors then apply supervised locality-preserving projections (SLPP) with heat kernel weights. The authors' feature extraction approach achieves a higher recognition rate compared to both traditional unsupervised LPP and SLPP with constant weights. For classification, the authors use the recently proposed sparse representation-based classification (SRC). However, instead of performing SRC using the computationally expensive ℓ_1 minimisation, the authors propose performing SRC using fast matching pursuit, which considerably improves the classification performance. The authors' proposed framework achieves ~99% recognition rate using four benchmark face databases, significantly faster than related frameworks.

1 Introduction

Face recognition has been one of the most studied pattern recognition problems over the last few decades. The face recognition problem can be stated as follows: given still or video images of a scene, it is required to identify a person using a database of face images. Compared to other biometric features, face recognition is natural, non-intrusive, and can be performed using images taken at a distance [1]. Face recognition is applied in many areas including security, access control, and surveillance.

While humans are capable of performing face recognition effectively for a vast number of subjects, automatic face recognition suffers from many difficulties. First of all, the dimensions of the face images are large. Furthermore, there may be variations of face images due to illumination, pose variation, and facial expressions. The presence of occlusion may pose more difficulty.

The face recognition problem involves two tasks: feature extraction and classification. Since human face images are typically high dimensional, dimensionality reduction techniques are usually associated with feature extraction [2]. A face image of size 256×256 can be considered as a point in 65,536-dimensional space. However, since face images are similar in the overall structure, they do not randomly occupy the whole image space, and can be described by a much lower dimensional subspace [3]. Therefore, dimensionality reduction techniques have been employed in most face recognition algorithms, which does not only result in complexity reduction, but also improves classification [2]. Among the most commonly used dimensionality reduction techniques are the principal component analysis (PCA) [3], Fisher's linear discriminant (FLD) [4], and locality-preserving projections (LPP) [5]. While PCA preserves the global structure of the image space, and FLD preserves the discriminating information, they fail to discover the underlying structure if the face images lie on a non-linear submanifold hidden in the image space [6]. LPP preserves local neighbourhood information resulting in closer projections for closer images. However, unsupervised LPP is mostly used, which ignores information about classes, and therefore, the recognition rates are not high. Although it is possible to apply dimensionality reduction directly to the pixels of the face images as in many existing frameworks, it has been shown that the application of such techniques to Gabor features, extracted from the images, results in a significant improvement in classification problems [7].

After the feature extraction stage comes the classification stage. Various classification techniques have been employed, including nearest neighbour, nearest subspace, support vector machines, and neural networks. Recently, a classification method termed sparse representation-based classification (SRC) [8] has been developed based on compressed sensing (CS) [9, 10], and shown to outperform other related classification techniques. CS is a sampling technique that is capable of reconstructing sparse (or compressible) signals from samples collected at a much lower rate than the Nyquist rate. The idea of SRC, and its recent improvements [11, 12], is based on representing the sample to be classified as a linear combination of the training samples. Such representation results in a sparse vector, since only components corresponding to the same class are significant. In order to obtain such sparse vector, CS recovery techniques are utilised. Traditionally, ℓ_1 minimisation is utilised. However, it is computationally expensive and not suitable for most real-time applications. Therefore, other algorithms have been proposed to reduce its complexity without affecting the accuracy. Among the utilised algorithms are fast ℓ_1 algorithms, such as the homotopy method [13], and greedy recovery algorithms, such as orthogonal matching pursuit (OMP) [14], which show some improvement in complexity.

In this paper, we propose a robust and efficient face recognition framework to (i) tackle the inefficient feature extraction resulting from either applying dimensionality reduction techniques that are incapable of non-linear manifold learning or applying dimensionality reduction directly to the image pixels, and (ii) significantly reduce the high complexity and long reconstruction times of CS-based classification in order to suite real-time applications. First of all, we extract Gabor features from face images. Gabor feature extraction improves the recognition performance and the robustness to variations of illumination, expression, and pose. Then, we apply supervised locality-preserving projections (SLPP) for dimensionality reduction, in contrast to most related frameworks that use unsupervised LPP. In contrast to [15] which uses simple-minded constant weights, we use heat kernel weights for the SLPP, which improves the learning process, and accordingly, improves the recognition rate. For classification, we apply SRC previously proposed in [8]. In contrast to [8], which uses ℓ_1 minimisation for CS recovery, we

propose applying our fast matching pursuit (FMP) algorithm [16]. This results in very efficient, fast, and accurate recognition, suitable for real-time applications, compared to other recovery algorithms.

The rest of the paper is organised as follows. Section 2 discusses the related literature. We present the feature extraction stage of our framework in Section 3 and the classification stage in Section 4. We present the simulation results in Section 5 and conclude in Section 6.

2 Related work

We briefly discuss the most relevant related face recognition approaches. The Eigenface algorithm [3] applies PCA to the face recognition problem. While PCA aims for good representation of data regardless of classes, FLD aims for the best classification as in the Fisherface algorithm [4]. However, PCA and FLD fail to discover the underlying structure if the face images lie on a non-linear submanifold hidden in the image space [6]. Laplacianfaces utilise LPP to preserve the local structure of the image space. The manifold structure is modelled by an adjacency graph which describes the local structure of the image space [6]. Multi-dimensional orthogonal transformations have also been used to reduce the implementation complexity [17, 18].

The aforementioned algorithms apply dimensionality reduction directly to the pixels of a face image. However, it has been shown that the application of such techniques to Gabor features extracted from the image results in a significant improvement in classification problems. Such features are motivated by receptive fields of simple cells in the primary visual cortex [19, 20]. Gabor features are also robust to variations of illumination, expression, and pose [7]. Gabor features are used with PCA in [21], with enhanced FLD in [22], and with LPP in [23]. Supervised LPP is applied to Gabor features in [15]. However, it uses simple-minded constant weights for the neighbourhood graph, which ignore the information about distances between samples of the same class. Therefore, they do not improve the recognition rate.

For classification, the traditional nearest-neighbour classifier is utilised in all of the previous algorithms. In [8], SRC is proposed based on CS, resulting in an improved recognition rate. SRC utilises CS recovery techniques for classification. In [8], ℓ_1 minimisation using primal-dual algorithm [24] is utilised for CS recovery. Such algorithm is computationally expensive and is reported to take a few seconds per test image in the same work. Also, homotopy algorithms [13] are mentioned in the same work as a faster alternative without providing simulation results. Dimensionality reduction is applied directly to image pixels without extracting Gabor features.

Various face recognition frameworks have been developed in the literature aiming to improve the basic SRC algorithm. For instance, the superposed SRC (SSRC) algorithm was proposed in [11]. SSRC uses two dictionaries. The first dictionary is used for the class centroids, while the second dictionary is used to represent the sample-to-centroid differences. Likewise, the semi-supervised sparse representation-based classification (S3RC) uses two dictionaries to represent faces [12]. The first dictionary consists of the training samples of each person. The second is a variation dictionary representing linear nuisance variables such as lighting and accessories. A Gaussian mixture model is then used for prototype face estimation. However, such frameworks suffer from common drawbacks. Namely, dimensionality reduction is performed using PCA. No Gabor features are extracted. Moreover, ℓ_1 minimisation is used for CS reconstruction using homotopy. Such drawbacks result in a reduced recognition accuracy and increased time.

Aiming to reduce the computational complexity of ℓ_1 minimisation techniques, OMP [14] has been used [25]. However, this comes at the expense of a reduced reconstruction accuracy. Furthermore, the fact that OMP performs one selection per iteration results in a large number of iterations, and consequently, large time compared to the FMP algorithm which we use in this paper.

It is worth mentioning that other classification approaches have been used in the literature such as nuclear norm-based matrix regression [26, 27] and contour-based binary descriptors [28]. Such classification approaches achieve recognition rates higher than SRC. However, their execution time, despite being faster in SRC, is still a challenging constraint that limits their use in real-life applications [26].

We conclude that the main challenges that face existing works are: (i) the use of dimensionality reduction techniques that are incapable of non-linear manifold learning or the application of dimensionality reduction directly to the image pixels, which significantly reduces the efficiency of feature extraction. (ii) The adoption of computationally complex approaches (such as the ℓ_1 minimisation used in CS-based classification and nuclear norm-based matrix regression) that result in long execution times, which make such approaches not suitable for real-time applications. These challenges motivate the proposed face recognition framework as described in the next section.

3 Feature extraction of the proposed framework

In this paper, we present a fast and efficient framework for face recognition using CS. The proposed dual-stage framework is composed of a feature extraction stage followed by a classification stage as in related approaches. In this section, we describe the feature extraction stage of the proposed framework. First, Gabor features are extracted from face images, and the resulting features are augmented into a column vector. The use of Gabor features increases the overall accuracy due to its immunity to variations under which the images were taken. Next, SLPP using heat kernel is applied to the resulting vectors. Our proposed feature extraction approach achieves higher recognition rates compared to existing approaches that either use traditional unsupervised LPP or use SLPP with constant weights.

3.1 Gabor features

Among the major challenges in the face recognition problem are the variations in illumination, facial expressions, and poses. Attempting to overcome such challenges, it was found that local features in face images are more robust against such variations [7]. A spatial-frequency analysis was found to be desirable to extract such features. It is known that wavelet analysis provides good space-frequency localisation. In particular, Gabor functions provide the optimised resolution in both the spatial and frequency domains [20, 29]. Therefore, Gabor wavelets are proposed to be the optimal basis for extracting local features for pattern recognition [7]. This is due to the fact that the shapes of Gabor functions are similar to the receptive fields of simple cells in the primary visual cortex.

For a two-dimensional greyscale image, the Gabor features are the magnitudes of the two-dimensional convolution of the image with the family of Gabor wavelets. Gabor features are then downsampled by a factor ρ to reduce the space dimension, and normalised to zero mean and unit variance [22]. Subsequently, the resultant convolution with each wavelet is converted into a single column vector by augmenting the columns of the image, and the resultant columns themselves are then augmented into a single column vector which represents the Gabor features of the input image.

3.2 Supervised locality-preserving projections

Following the extraction of Gabor features, dimensionality reduction is applied. While Wright *et al*. [8] suggest that any dimensionality reduction technique, even multiplication by a random matrix, would result in high recognition rate, given that a sufficient number of features is used, we suggest that it is better to select dimensionality reduction techniques that require the least number of features for the same recognition rate, or alternatively, those that achieve the highest recognition rate for the same number of features. Using fewer number of features results in faster recognition, and consequently results in a more efficient algorithm as we shall demonstrate using simulations.

Input: A matrix of training samples

$$\boldsymbol{A} = [\boldsymbol{A}_1, \boldsymbol{A}_2, \ldots, \boldsymbol{A}_{nc}] \in \mathbb{R}^{\mathrm{k}\times \mathrm{n}}$$

for *nc* classes, a test sample $\boldsymbol{y} \in \mathbb{R}^{\mathrm{k}}$.

1. Normalize the columns of $\boldsymbol{A}$ to have unit ℓ_2-norm.

2. Solve $\boldsymbol{y} = \boldsymbol{A}\boldsymbol{x}$ using FMP (given in the following section).

3. Compute the residuals

$$r_i(\boldsymbol{y}) = \|\boldsymbol{y} - \boldsymbol{A}\delta_i(\hat{\boldsymbol{x}})\|_2$$

for $i = 1, \ldots, nc$, where $\delta_i(\hat{\boldsymbol{x}})$ is a new vector whose only nonzero entries are the entries in $\hat{\boldsymbol{x}}$ corresponding to class i.

Output: $\text{Identity}(\boldsymbol{y}) = \arg\min_i r_i(\boldsymbol{y})$

Fig. 1 *Algorithm 1: Sparse representation-based classification (SRC)*

3.2.1 Manifold learning: Manifold learning techniques preserve the local structure of the face space [6]. In other words, points that are closer in the higher-dimensional face space are mapped to closer points in the low-dimensional space, and vice versa. A manifold is a subset of a Euclidean space $\mathbb{R}^{\mathrm{n}}$ that is locally Euclidean of dimension $m < n$. For example, a curve drawn on a piece of paper belongs to $\mathbb{R}^2$; however, locally, it is like a line belonging to $\mathbb{R}^1$, therefore, it is a one-dimensional manifold. It was proposed that variations in face images of a single person, due to different poses and expressions, lie on a non-linear submanifold of the face space [30, 31].

Our framework uses SLPP for manifold learning. Contrary to most algorithms that use the traditional unsupervised LPP, we use supervised LPP, which takes classes of samples in account when forming the graph. Furthermore, we use a heat kernel for weights to improve the learning process, which in turn improves the recognition rate.

The manifold structure is modelled by a graph which describes the local structure of the image space [6]. Each training sample is represented by a vertex in the graph. For traditional unsupervised LPP, a nearest neighbour graph is used, where two vertices are connected if either one lies in the k nearest neighbours of the other. For supervised LPP, two vertices are connected if they belong to the same class. This improves the classification ability of the projection. Furthermore, we use the heat kernels for graph weights, where the weight of the edge connecting the vertices $\boldsymbol{x}_i$ and $\boldsymbol{x}_j$ is $\mathrm{e}^{-\|\boldsymbol{x}_i - \boldsymbol{x}_j\|^2/t}$ [32], where $t \in \mathbb{R}$ determines the rate of decay. Therefore, the weights of the edges of the graph are given by:

$$S_{ij} = \begin{cases} \mathrm{e}^{-\|\boldsymbol{x}_i - \boldsymbol{x}_j\|^2/t} & \text{if } x_i \text{ and } x_j \text{ belongtothesameclass} \\ 0 & \text{otherwise} \end{cases} \quad (1)$$

Using such weights, we proceed as in the Laplacianface algorithm [6] to perform dimensionality reduction.

4 FMP-based classification

In this section, we describe the classification stage of our proposed framework. We use the SRC algorithm proposed in [8]. However, instead of using ℓ_1 minimisation for CS recovery, we use our FMP algorithm. This results in a significant speedup with very close recognition rate. SRC is based on the theory of CS. We first review CS, then SRC. Finally, we show how to apply FMP for classification.

4.1 CS preliminaries

Consider a sparse signal $\boldsymbol{x} \in \mathbb{R}^n$, of sparsity k. A measurement system that acquires m linear measurements, where $m \ll n$, can be represented as

$$\boldsymbol{y} = \boldsymbol{\Phi}\boldsymbol{x}, \quad (2)$$

where $\boldsymbol{\Phi} \in \mathbb{R}^{m\times n}$ is the sensing or measurement matrix, and $\boldsymbol{y} \in \mathbb{R}^m$ the measurement vector or the samples. The original signal $\boldsymbol{x}$ can be recovered from the measurement vector $\boldsymbol{y}$, provided that the sensing matrix satisfies the restricted isometry property [33]. This applies to matrices of entries that are independent and identically distributed and follow a Gaussian, Bernoulli, or sub-Gaussian distribution with high probability. Donoho [10] originally suggested using ℓ_1 minimisation to reconstruct the sparse signal as follows:

$$\hat{\boldsymbol{x}} = \arg\min_z \ \| z \|_1 \ \text{subjectto} \ \boldsymbol{y} = \boldsymbol{\Phi}\boldsymbol{z} \quad (3)$$

4.2 SRC overview

Based on the CS theory, SRC was proposed and found to achieve better recognition rates compared to traditional classifiers such as nearest neighbour [8]. Here, we review the SRC algorithm.

Suppose a training set contains samples from *nc* subjects, and a test sample $\boldsymbol{y}$ is to be classified. Assume that the test sample can be approximately represented by a linear combination of training samples associated with the corresponding subject. Now let $\boldsymbol{A}_i$ be a matrix whose columns are the training samples of subject i. Construct the augmented matrix $\boldsymbol{A}$ as:

$$\boldsymbol{A} = [\boldsymbol{A}_1, \boldsymbol{A}_2, \cdots, \boldsymbol{A}_{nc}] \quad (4)$$

Assume that the vector $\boldsymbol{x}$ contains the weights required to form the test sample $\boldsymbol{y}$ from the training samples. Therefore, $\boldsymbol{y}$ can be expressed as

$$\boldsymbol{y} = \boldsymbol{A}\boldsymbol{x} = [\boldsymbol{A}_1, \boldsymbol{A}_2, \cdots, \boldsymbol{A}_{nc}]\boldsymbol{x} \quad (5)$$

The vector $\boldsymbol{x}$ contains larger-magnitude components at indices corresponding to the subject that the test sample belongs to, and very small or zero components at other indices. Therefore, the vector $\boldsymbol{x}$ is sparse and can be reconstructed using CS techniques. Let the reconstructed vector be $\hat{\boldsymbol{x}}$. Then, $\boldsymbol{y}$ is reconstructed from linear combinations of the columns of $\boldsymbol{A}$ using indices of $\hat{\boldsymbol{x}}$ associated with each subject individually. The sample is then classified as belonging to the subject corresponding to the least reconstruction error (see Fig. 1).

4.3 Proposed FMP

FMP is a fast and accurate recovery algorithm for CS that is suitable for operation with data of larger sizes. In order to identify the support of the sparse signal $\boldsymbol{x}$ (its non-zero indices), FMP correlates $\boldsymbol{y}$ with the columns of the sensing matrix. In this algorithm, the sensing matrix $\boldsymbol{\Phi}$ will be taken as the matrix $\boldsymbol{A}$ given by (4). This contrasts with [16] in which we applied the FMP concepts to reconstruct a wideband frequency spectrum from time domain samples for cognitive radio applications. The FMP algorithm then iteratively identifies the support of the sparse signal by adaptively selecting elements from a reduced set of the correlation values. The signal is then estimated based on this support, through least square minimisation. In FMP, least square minimisation is performed iteratively avoiding large matrix inversion, which results in significant complexity reduction. Instead of forming the pseudo-inverse required for least square minimisation directly with matrix inversion, FMP calculates it in each iteration from data in the previous iteration. FMP then prunes the signal estimate to exclude the incorrectly selected elements. Then a residual is calculated and the aforementioned steps are repeated until a stopping condition is met. FMP is summarised in Algorithm 2 (see Fig. 2) and its components are explained in what follows.

4.3.1 Support identification: Selection in the FMP algorithm is based on a double-thresholding technique [34]. This selection strategy adaptively selects elements from a reduced set of the correlation vector. First, the elements from which we perform selection are reduced to a set containing the top magnitude elements (using a certain threshold β). Then, elements of larger magnitude than a fraction $0 < \alpha < 1$ of the maximum element are selected from the reduced set, and their indices are added to the support set. Properly selecting α and β leads to the selection of an optimum number of elements per iteration. Using exhaustive

Input: Sensing matrix $\boldsymbol{\Phi}$ (matrix containing training samples), measurement vector $\boldsymbol{y}$ (sample to be classified), sparsity k (number of samples per class), parameters α and β.

Initialise: $\hat{\boldsymbol{x}}^{[0]} = 0, r^{[0]} = \boldsymbol{y}, T^{[0]} = \emptyset$.

for $i = 1; i := i + 1$ until the stopping criterion is met **do**

$g^{[i]} \leftarrow \boldsymbol{\Phi}^* r^{[i-1]}$ {Form signal proxy}

$J \leftarrow L_{\beta k}(g^{[i]})$ {Indices of βk largest magnitude elements in g}

$W \leftarrow \{j : g_j^{[i]} \geq \alpha \max l|g_l^{[i]}|, j \in J\}$ {indices of elements in J larger than or equal to $\alpha \max l|g_l^{[i]}|$}

$U \leftarrow W \setminus T$ {Indices of newly added atoms not previously identified}

$T \leftarrow (T\ U)$ {Support merging}

if i==1 **then**

$\boldsymbol{E} \leftarrow (\boldsymbol{\Phi}_U^{\mathrm{T}} \boldsymbol{\Phi}_U)^{-1}$

$\boldsymbol{\Phi}^{\dagger} \leftarrow \boldsymbol{E}\boldsymbol{\Phi}_U^{T}$

else

if $|\boldsymbol{D} - \boldsymbol{C}\boldsymbol{A}^{-1}\boldsymbol{B}|$==0

$U \leftarrow U_1$ {if $\boldsymbol{S} = \boldsymbol{D} - \boldsymbol{C}\boldsymbol{A}^{-1}\boldsymbol{B}$ is singular, only use the first element in U}

end if

$\boldsymbol{Q} \leftarrow \boldsymbol{\Phi}_U$

$\boldsymbol{A}^{-1} \leftarrow \boldsymbol{E}$, $\boldsymbol{B} \leftarrow \boldsymbol{\Phi}_T^{\mathrm{T}}\boldsymbol{Q}$, $\boldsymbol{C} \leftarrow \boldsymbol{Q}^{\mathrm{T}}\boldsymbol{\Phi}_T$, $\boldsymbol{D} \leftarrow \boldsymbol{Q}^{\mathrm{T}}\boldsymbol{Q}$

$\boldsymbol{S} \leftarrow \boldsymbol{D} - \boldsymbol{C}\boldsymbol{A}^{-1}\boldsymbol{B}$

$\boldsymbol{E} \leftarrow \begin{pmatrix} \boldsymbol{A}^{-1} + \boldsymbol{A}^{-1}\boldsymbol{B}\boldsymbol{S}^{-1}\boldsymbol{C}\boldsymbol{A}^{-1} & -\boldsymbol{A}^{-1}\boldsymbol{B}\boldsymbol{S}^{-1} \\ -\boldsymbol{S}^{-1}\boldsymbol{C}\boldsymbol{A}^{-1} & \boldsymbol{S}^{-1} \end{pmatrix}$ $\{\boldsymbol{E} = (\boldsymbol{\Phi}_T^{\mathrm{T}}\boldsymbol{\Phi}_T)^{-1}\}$

$\boldsymbol{\Phi}_T^{\dagger} \leftarrow \boldsymbol{E}\boldsymbol{\Phi}_T^{\mathrm{T}}$

end if

$b|_T \leftarrow \boldsymbol{\Phi}_T^{\dagger} y$, $b|_{T^c} \leftarrow 0$ {Signal estimation}

$\hat{\boldsymbol{x}}^{[i]} \leftarrow H_k(b)$ {Prune approximation}

$V \leftarrow T \setminus L_k(b)$ {Pruned indices}

if $|V| > 0$ **then**

$T \leftarrow T \setminus V$ {Update support set by removing pruned indices}

$\boldsymbol{R} \leftarrow \boldsymbol{\Phi}_V$ {Columns to be pruned}

$t \leftarrow |T|$, $v \leftarrow |V|$ {Cardinality of sets T and V }

$\begin{pmatrix} \boldsymbol{A}_{t-v\times t-v} & \boldsymbol{B}_{t-v\times v} \\ \boldsymbol{C}_{v\times t-v} & \boldsymbol{D}_{v\times v} \end{pmatrix} \leftarrow \boldsymbol{E}$ {Partition $\boldsymbol{E}$ into four matrices}

$\boldsymbol{S} \leftarrow \boldsymbol{D} - \boldsymbol{C}\boldsymbol{A}^{-1}\boldsymbol{B}$

$\boldsymbol{N}_{11} \leftarrow \boldsymbol{A}^{-1} + \boldsymbol{A}^{-1}\boldsymbol{B}\boldsymbol{S}^{-1}\boldsymbol{C}\boldsymbol{A}^{-1}$, $\boldsymbol{N}_{12} \leftarrow -\boldsymbol{A}^{-1}\boldsymbol{B}\boldsymbol{S}^{-1}$, $\boldsymbol{N}_{21} \leftarrow -\boldsymbol{S}^{-1}\boldsymbol{C}\boldsymbol{A}^{-1}$, $\boldsymbol{N}_{22} \leftarrow \boldsymbol{S}^{-1}$

$\boldsymbol{E} \leftarrow \boldsymbol{N}_{11} - \boldsymbol{N}_{12}\boldsymbol{N}_{22}^{-1}\boldsymbol{N}_{21}$ {Update $\boldsymbol{E}$ removing pruned columns}

end if

$\boldsymbol{r} \leftarrow \boldsymbol{y} - \boldsymbol{\Phi}\hat{\boldsymbol{x}}^{[i]}$ {Update residual}

end for

Output: Sparse representation $\hat{\boldsymbol{x}}$

Fig. 2 *Algorithm 2: Fast matching pursuit*

simulation of α and β values, we found that moderate values of both $\alpha \in [0.5, 0.7]$ and $\beta \in [0.15, 0.75]$ result in the best performance because the number of selected elements per iteration becomes neither too large nor too small. Furthermore, the simulation results indicated that the algorithm performance is not sensitive to the specific α and β values as long as they are in the aforementioned optimum range. More details regarding the values of α and β are available in [35].

4.3.2 Signal estimation: Based on the identified support set, a signal estimate $\hat{\boldsymbol{x}}$ is obtained through least square minimisation. Let $\boldsymbol{\Phi}_i$ denote a matrix containing the columns of the sensing matrix at indices from the identified support set in the ith iteration. Least square minimisation is typically performed by multiplication by the pseudo-inverse of $\boldsymbol{\Phi}_i$, given by $\boldsymbol{\Phi}_i^{\dagger} = (\boldsymbol{\Phi}_i^{\mathrm{T}}\boldsymbol{\Phi}_i)^{-1}\boldsymbol{\Phi}_i^{\mathrm{T}}$. However, in FMP, we avoid performing the matrix inversion. Instead, we obtain $\boldsymbol{\Phi}_i^{\dagger}$ using $(\boldsymbol{\Phi}_{i-1}^{\mathrm{T}}\boldsymbol{\Phi}_{i-1})^{-1}$ from the previous iteration. This is done in two steps. First, the inverse from the previous iteration is updated by adding the newly selected columns to the matrix $\boldsymbol{\Phi}_{i-1}$. Then, after pruning, it is updated by removing the pruned columns.

Let $\boldsymbol{Q}$ be a matrix containing the newly added columns. Augmenting the matrix $\boldsymbol{\Phi}_{i-1}$ with $\boldsymbol{Q}$, we have:

$$\boldsymbol{\Phi}_i = (\boldsymbol{\Phi}_{i-1}\boldsymbol{Q}) \tag{6}$$

Applying the Schur–Banachiewicz inverse formula [36] to iteratively obtain $(\boldsymbol{\Phi}_i^{\mathrm{T}}\boldsymbol{\Phi}_i)^{-1}$ without inverting the matrix $\boldsymbol{\Phi}_i^{\mathrm{T}}\boldsymbol{\Phi}_i$. Thus, we obtain

$$(\boldsymbol{\Phi}_i^{\mathrm{T}}\boldsymbol{\Phi}_i)^{-1} = \begin{pmatrix} \boldsymbol{A}^{-1} + \boldsymbol{A}^{-1}\boldsymbol{B}\boldsymbol{S}^{-1}\boldsymbol{C}\boldsymbol{A}^{-1} & -\boldsymbol{A}^{-1}\boldsymbol{B}\boldsymbol{S}^{-1} \\ -\boldsymbol{S}^{-1}\boldsymbol{C}\boldsymbol{A}^{-1} & \boldsymbol{S}^{-1} \end{pmatrix} \tag{7}$$

where $\boldsymbol{A} = \boldsymbol{\Phi}_{i-1}^{\mathrm{T}}\boldsymbol{\Phi}_{i-1}, \boldsymbol{B} = \boldsymbol{\Phi}_{i-1}^{\mathrm{T}}\boldsymbol{Q}, \boldsymbol{C} = \boldsymbol{Q}^{\mathrm{T}}\boldsymbol{\Phi}_{i-1}, \boldsymbol{D} = \boldsymbol{Q}^{\mathrm{T}}\boldsymbol{Q}$. The Schur complement, $\boldsymbol{S} = \boldsymbol{D} - \boldsymbol{C}\boldsymbol{A}^{-1}\boldsymbol{B}$, is a matrix of dimensions equal to the number of added columns, which is typically small. Matrix inversion is performed only in the first iteration to obtain $(\boldsymbol{\Phi}_1^{\mathrm{T}}\boldsymbol{\Phi}_1)^{-1}$.

4.3.3 Pruning: The estimated signal is pruned, keeping only the top k magnitude components of the estimated signal and setting the rest to zero. Thus, the support set is updated by removing the elements with the least contribution to the estimated signal. This improves the reconstruction accuracy and speed. The inverse $(\boldsymbol{\Phi}_i^{\mathrm{T}}\boldsymbol{\Phi}_i)^{-1}$ is then updated removing the columns corresponding to the pruned elements. Let us denote the sensing matrix in the ith iteration after removing the pruned columns by $\boldsymbol{\Phi}_{i+0.5}$, and the matrix consisting of the columns to be pruned by $\boldsymbol{R}$, which are to removed from $\boldsymbol{\Phi}_i$ to obtain $\boldsymbol{\Phi}_{i+1}$. The matrix $\boldsymbol{\Phi}_i$ can be expressed as

$$\boldsymbol{\Phi}_i = (\boldsymbol{\Phi}_{i+0.5}\ \boldsymbol{R}) \tag{8}$$

The inverse $(\boldsymbol{\Phi}_i^{\mathrm{T}}\boldsymbol{\Phi}_i)^{-1}$ is performed using the Schur–Banachiewicz inverse formula as in the previous step, but this time in the reverse manner. The inverse of a smaller matrix is obtained using the inverse of a larger matrix. We have the inverse of a matrix $\boldsymbol{M} = \boldsymbol{\Phi}_i^{\mathrm{T}}\boldsymbol{\Phi}_i$ and we want to obtain the inverse of a matrix $\boldsymbol{A} = \boldsymbol{\Phi}_{i+0.5}^{\mathrm{T}}\boldsymbol{\Phi}_{i+0.5}$, such that $\boldsymbol{\Phi}_i = (\boldsymbol{\Phi}_{i+0.5}\ \boldsymbol{R})$. Let us define the matrix $\boldsymbol{N}$ as follows:

$$\boldsymbol{N} = \boldsymbol{M}^{-1} = (\boldsymbol{\Phi}_i^{\mathrm{T}}\boldsymbol{\Phi}_i)^{-1} = \begin{pmatrix} \boldsymbol{\Phi}_{i+0.5}^{\mathrm{T}}\boldsymbol{\Phi}_{i+0.5} & \boldsymbol{\Phi}_{i+0.5}^{\mathrm{T}}\boldsymbol{R} \\ \boldsymbol{R}^{\mathrm{T}}\boldsymbol{\Phi}_{i+0.5} & \boldsymbol{R}^{\mathrm{T}}\boldsymbol{R} \end{pmatrix}^{-1} \tag{9}$$

Taking $\boldsymbol{A} = \boldsymbol{\Phi}_{i+0.5}^{\mathrm{T}}\boldsymbol{\Phi}_{i+0.5}, \boldsymbol{B} = \boldsymbol{\Phi}_{i+0.5}^{\mathrm{T}}\boldsymbol{R}, \boldsymbol{C} = \boldsymbol{R}^{\mathrm{T}}\boldsymbol{\Phi}_{i+0.5}$, and $\boldsymbol{D} = \boldsymbol{R}^{\mathrm{T}}\boldsymbol{R}$, we have:

$$\boldsymbol{N} = \begin{pmatrix} \boldsymbol{A} & \boldsymbol{B} \\ \boldsymbol{C} & \boldsymbol{D} \end{pmatrix}^{-1} = \begin{pmatrix} \boldsymbol{N}_{11} & \boldsymbol{N}_{12} \\ \boldsymbol{N}_{21} & \boldsymbol{N}_{22} \end{pmatrix} \tag{10}$$

where
$\boldsymbol{N}_{11} = \boldsymbol{A}^{-1} + \boldsymbol{A}^{-1}\boldsymbol{B}\boldsymbol{S}^{-1}\boldsymbol{C}\boldsymbol{A}^{-1}, \boldsymbol{N}_{12} = -\boldsymbol{A}^{-1}\boldsymbol{B}\boldsymbol{S}^{-1}, \boldsymbol{N}_{21} = -\boldsymbol{S}^{-1}\boldsymbol{C}\boldsymbol{A}^{-1},$,
$\boldsymbol{N}_{22} = \boldsymbol{S}^{-1}$
from which we can obtain $\boldsymbol{A}^{-1}$ as:

$$\boldsymbol{A}^{-1} = (\boldsymbol{\Phi}_{i+0.5}^{\mathrm{T}}\boldsymbol{\Phi}_{i+0.5})^{-1} = \boldsymbol{N}_{11} - \boldsymbol{N}_{12}\boldsymbol{N}_{22}^{-1}\boldsymbol{N}_{21} \tag{11}$$

4.3.4 Residual calculation: A residual is then calculated by subtracting the contribution of the estimated signal after pruning from $\boldsymbol{y}$. The residual is given by $\boldsymbol{r} = \boldsymbol{y} - \boldsymbol{\Phi}\hat{\boldsymbol{x}}$.

We summarise the proposed algorithm in Algorithm 2 (see Fig. 2). The operator $L_k(\cdot)$ returns the index set of the k largest absolute values of the elements of its argument vector. The hard thresholding operator $H_k(\cdot)$ retains only the k elements with the largest absolute values and sets the rest to zero.

5 Experimental evaluation

First, we introduce the databases used. Then, we describe the simulation set-up. Finally, we present the simulation results.

5.1 Used databases

We use the following four databases for our performance evaluation. In what follows, we briefly highlight the characteristics of the images included in each database. No illumination correction is performed to any of the used databases.

5.1.1 AR face database: The AR face database [37] contains over 4000 coloured images corresponding to 126 people's faces. No restrictions on wear (clothes, glasses etc.), make-up, hair style

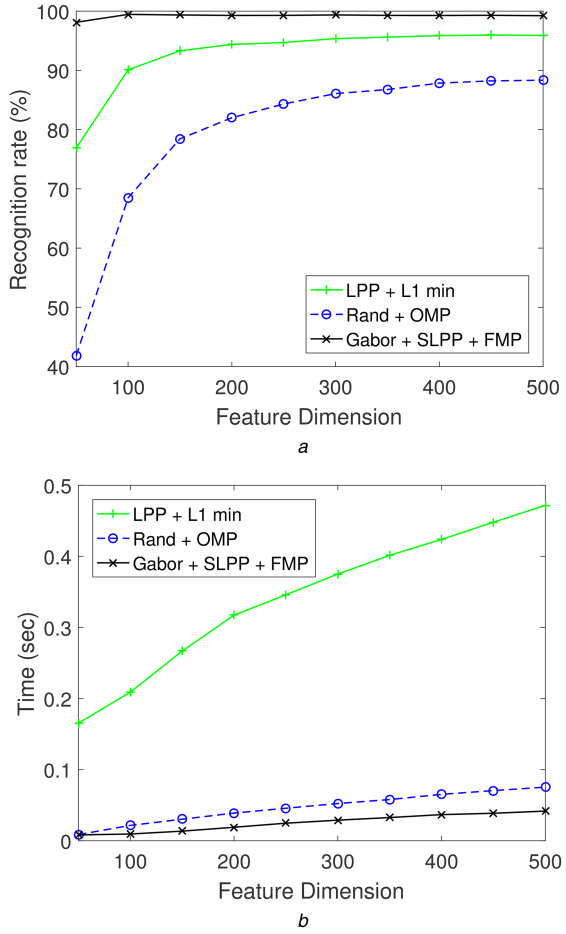

Fig. 3 *Recognition rate and time compared to the basic algorithms*
(a) Recognition rate, ***(b)*** Recognition time

etc. were imposed to participants. We used the subset developed in [38] which contain the images of 100 persons (50 men and 50 women), with 26 images per subject. The images of this subset are cropped to 120×165 pixels and then converted to greyscale and resized to 128×128 pixels [38].

5.1.2 Extended Yale face database B: Extended Yale face database B [39] contains facial images of 38 persons, with 64 images per person under varying pose and illumination. The images included in this database are cropped to 168×192 pixels.

5.1.3 ORL database: This database contains 10 different images of each of 40 persons. For some subjects, the images were taken at different times, varying the lighting, facial expressions, with or without glasses. The ORL database images are of size 92×112 pixels.

5.1.4 CAS-PEAL-R1 database: CAS-PEAL-R1 face database consists of 30,900 images of 1040 subjects. Face images have different variations, including pose, expression, accessories, and lighting. We use a subset of containing images of 100 persons, with 14 images per subject. The images included in this database are cropped to 190×290 pixels.

5.2 Simulation set-up

We adopted the widely used cross-validation approach for performance evaluation. However, the image set is divided into ten subsets in typical 10-fold cross-validation: nine subsets are used for training and one subset is used for testing. This implies that few images are used for testing which is not the case in real-life scenarios. Therefore, we used one half of the images that are randomly selected for training in each iteration, and the other half is used for testing. This implies that the training set contains a reasonable number of images, and at the same time, enough testing images are available. The presented results are averaged over ten iterations. Gabor wavelets of five scales and eight orientations are used. The features are then downsampled by a factor $\rho = 64$. Dimensionality reduction is applied to the resultant features and then classification is performed using the SRC algorithm. For FMP, we take $\alpha = 0.7$ and $\beta = 0.25$, even though the FMP accuracy is insensitive to the α and β values [35]. For the homotopy algorithm, we follow Yang *et al*. [40]. In our experiments, we implemented the proposed framework and the benchmark approaches using Matlab running on a computer with a 64-bit Intel® Pentium® 2.8 GHz processor with 6 GB RAM.

5.3 Simulation results

In this section, we present detailed simulation results using the AR face database and brief results using other databases. Our proposed algorithm has three main components: (i) Gabor feature extraction, (ii) SLPP for dimensionality reduction, and (iii) FMP for SRC. First, we compare the overall performance of our algorithm against other related algorithms. Then, we evaluate the performance of each component of our algorithm by comparing it to corresponding components in the related algorithms. In doing so, we perform three sets of experiments. We fix two stages and compare the third stage against related approaches. First, we test SRC using FMP against other CS reconstruction techniques. Then, we test the dimensionality reduction using SLPP against other techniques. Then, we evaluate the benefit from using Gabor features. We also present brief simulation results using other databases. Finally, we compare the overall performance of our framework against other frameworks that are improvements to the basic SRC algorithm.

5.3.1 Comparison with basic algorithms: We compare our proposed algorithm against the related algorithms proposed in [8, 25]. In [8], various features were extracted from face images. The best recognition rates were obtained for Laplacianfaces and randomfaces (random projections). For classification, SRC was proposed, and ℓ_1 minimisation was used for CS recovery. On the other hand, Vo *et al*. [25] used randomfaces as features, and used OMP for CS recovery in SRC. In both algorithms, dimensionality reduction was directly applied to the image pixels without taking Gabor features.

Fig. 3 depicts the recognition rate and the SRC reconstruction time. As expected, as the feature dimension increases, the recognition rate improves, while the reconstruction time increases. Our algorithm achieves the highest recognition rate of ~99.45% at 100 features taking the lowest time of ~10 ms. Laplacianfaces with ℓ_1 minimisation [8] yields significantly less recognition rate of 95.35% at 300 features taking 380 ms. Increasing the feature dimension beyond 300 shows little improvement. Randomfaces with OMP [25] results in a recognition rate of 88.22% at 450 features taking 70 ms.

5.3.2 Comparison of CS recovery techniques used for SRC: Next, we compare each component of our algorithm separately against corresponding components in other algorithms. We show the benefit of using FMP for classification based on sparse representation. We present the results of ℓ_1 minimisation (using homotopy), OMP, and FMP. We plot the recognition rate and time versus the number of features used. It is reported in [8] that ℓ_1 minimisation using primal-dual algorithm takes a few seconds per image, which is consistent with our simulations. Wright *et al*. [8] also suggest that homotopy may be used for faster operation. Our simulations show that both achieve very close recognition rates; therefore, we only present the results of homotopy algorithm which is more efficient.

We use our proposed heat kernel SLPP applied to Gabor features, and perform SRC using different CS reconstruction algorithms. Fig. 4 illustrates the recognition rate and time against the number of features used. All reconstruction algorithms achieve very high and very close recognition rates that exceed 99%, due to the use of Gabor features together with SLPP. Specifically, at a feature dimension of 100, the reconstruction rate is ~99.4%.

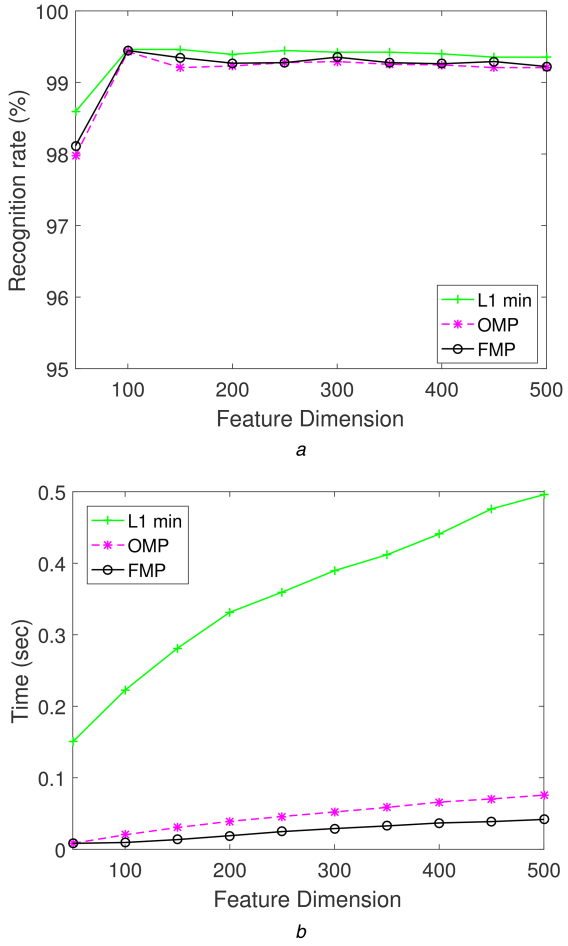

Fig. 4 *Recognition rate and time using different CS recovery algorithms*
(a) Recognition rate, ***(b)*** Recognition time

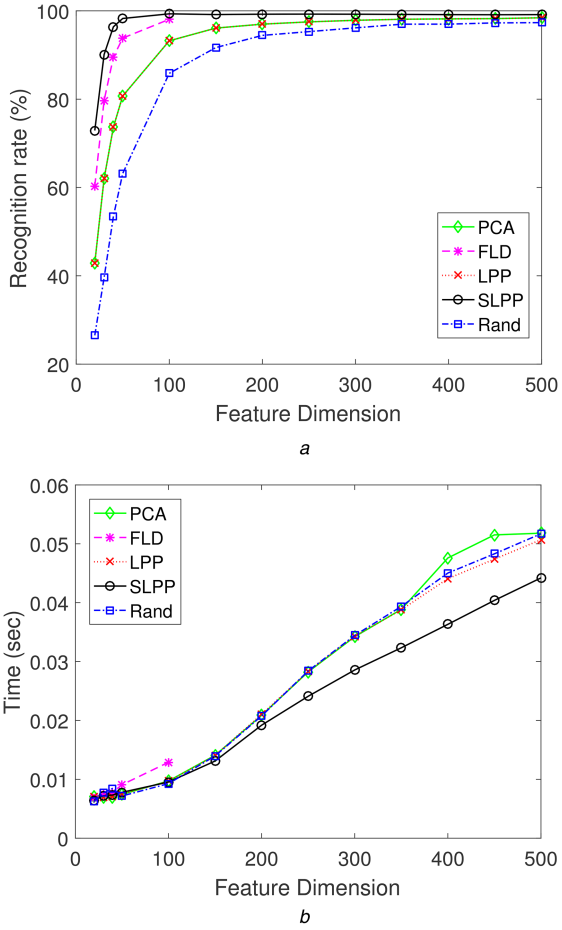

Fig. 5 *Recognition rate and time using different dimensionality reduction techniques applied to Gabor features*
(a) Recognition rate, ***(b)*** Recognition time

However, using FMP takes the least reconstruction time of ∼0.01 s at 100 features, compared to 0.02 s for OMP, and 0.22 for ℓ_1. It can be seen that as the dimension increases, the reconstruction time increases rapidly for ℓ_1 minimisation, which reaches ∼0.5 s for 500 features. OMP takes 0.08 s, while FMP takes 0.04 s at that point. This shows that FMP is capable of accurate reconstruction, considerably faster than other algorithms.

5.3.3 Comparison of dimensionality reduction techniques: We next compare the performance of different dimensionality reduction techniques. Fig. 5 depicts the recognition rate and time using different dimensionality reduction techniques applied to Gabor features. SRC algorithm is used for classification, applying FMP. SLPP achieves the highest recognition rate of 99.4% using 100 features. Next comes FLD, LPP, PCA, and finally random projections [8]. While it is possible to increase the dimension to obtain higher recognition rates for other dimensionality reduction techniques as argued in [8], this comes at the expense of increased reconstruction time. At 100 features, SLPP takes ∼0.01 s to achieve a recognition rate 99.4%. Random projections take ∼0.05 s achieving a recognition rate of 97.4% at 500 features. Therefore, SLPP achieves better recognition rates using smaller number of features, less reconstruction time, compared to other techniques. While increasing the number of features increases the recognition rate for other techniques, this comes at the expense of increased reconstruction time. It is noted that FLD uses a maximum of 99 features, which is one less than the number of classes.

5.3.4 Effect of Gabor features: In the following experiment, we demonstrate the benefit from using Gabor features. While the AR database contains images with large variations in pose, illumination, and occlusion, our proposed framework achieves >99% recognition rate as illustrated in Fig. 5. We now repeat the comparison of different dimensionality reduction techniques applied directly to the image pixels without extracting Gabor features. As can be seen in Fig. 6, the reconstruction rates decrease significantly. A maximum recognition rate of ∼92.5% is obtained through SLPP at 500 features. This takes ∼0.047 s.

5.3.5 Other databases: In this section, we present the results of our simulations using the extended Yale database, the ORL AT&T, and the CAS-PEAL-R1 databases. We only show the feature dimensions that correspond to the highest recognition rates obtained for each technique, or after which the improvement in recognition rate becomes insignificant. We also add data from our simulations of the AR database corresponding to the above figures. Table 1 compares different CS recovery techniques used for SRC, using SLPP applied to Gabor features. Our proposed framework achieves higher recognition rates at significantly lower times using all the databases that we used for performance evaluation.

5.3.6 Comparison with recent SRC improvements: In all of the above experiments, we evaluated the contributions of the individual components of the proposed framework (Gabor feature extraction, SLPP, and FMP) in the overall performance. In this experiment, we compare the overall performance of our SRC-based framework against two recent improvements of the basic SRC algorithm, namely, the SSRC [11] and the semi-supervised SRC (S3RC) [12]. We only consider the AR face database and the more challenging CAS-PEAL-R1 database. Fig. 7 illustrates the recognition rate and time using the CAS-PEAL-R1 database. Our proposed framework results in a significant improvement in the recognition rate compared to the two improved SRC frameworks. This is due to our feature selection that is based on Gabor and SLPP using heat kernel. Furthermore, our algorithm achieves a significant speed-up compared to the other algorithms. This is due to the use of the fast and efficient FMP algorithm, in contrast to the

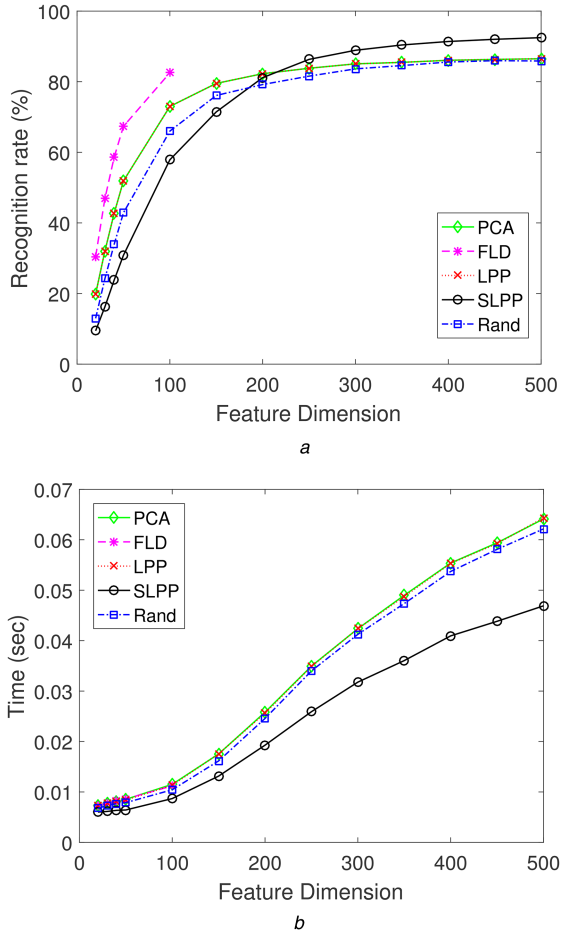

Fig. 6 *Recognition rate and time using different dimensionality reduction techniques without Gabor features*
(a) Recognition rate, ***(b)*** Recognition time

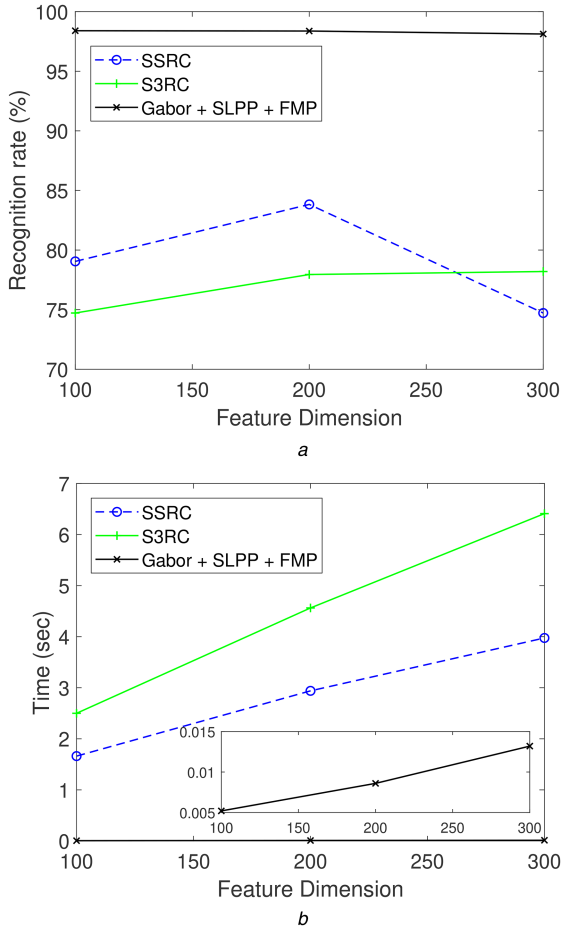

Fig. 7 *Recognition rate and time of the different SRC-based face recognition techniques*
(a) Recognition rate, ***(b)*** Recognition time

Table 1 Summary of best feature dimensions, recognition rate, and reconstruction times

Database	Rec. Algo.	Features	Rec., %	Time, ms
AR	LPP + ℓ_1 min	200	94.38	317
	Rand + OMP	300	86.1	52
	Gabor + SLPP + FMP	**100**	**99.45**	**10**
Ext. Yale	LPP + ℓ_1 min	200	93.45	732
	Rand + OMP	250	95.14	118
	Gabor + SLPP + FMP	**50**	**98.9**	**16**
ORL	LPP + ℓ_1 min	100	93.45	136
	Rand + OMP	150	88.7	2
	Gabor + SLPP + FMP	**90**	**99.3**	**2**
CAS-PEAL-R1	LPP + ℓ_1 min	300	81.03	333.2
	Rand + OMP	200	72.33	23.3
	Gabor + SLPP + FMP	**130**	**98.68**	**6**

other ℓ_1-based algorithms such as homotopy. Table 2 summarises our experiments on the AR and CAS-PEAL-R1 databases, given the best feature dimension, and the corresponding recognition rate and time. Again, our algorithm achieves significant improvement in both the reconstruction rate and time.

6 Conclusion

In this paper, we have presented a CS-based framework for efficient, yet fast, face recognition. The feature extraction component of the framework extracts Gabor features and then applies SLPP with heat kernel. This provides robustness against many sorts of variations in face images. For classification, the proposed framework uses FMP for SRC. This significantly improves the speed and efficiency of our framework compared to other SRC schemes.

Table 2 Summary of best feature dimensions, recognition rate, and reconstruction times against recent SRC variations

Database	Rec. Algo.	Features	Rec., %	Time, ms
AR	SSRC [11]	300	98.04	2718
	S3RC [12]	200	92.45	3665
	Gabor + SLPP + FMP	**100**	**99.45**	**10**
CAS-PEAL-R1	SSRC [11]	300	86.10	3974
	S3RC [12]	200	77.95	4562
	Gabor + SLPP + FMP	**130**	**98.68**	**6**

Simulation results have shown that the proposed framework achieves a recognition rate of almost 99% using four benchmark face databases, significantly faster than other frameworks. We also compared each component of our framework to other related components in order to assess the contribution of each component in the overall performance. FMP achieves a significant speedup compared to other CS recovery algorithms, including ℓ_1 minimisation and OMP. The extraction of Gabor features significantly improves the recognition rate. SLPP with heat kernel achieves the best recognition rate compared to other dimensionality reduction techniques. Compared to other frameworks that are based on the SRC algorithm, we have shown that our framework shows significant improvements in the recognition rate and time.

7 References

[1] Jain, A.K., Li, S.Z.: '*Handbook of face recognition*' (Springer, New York, 2011)

[2] Sun, B., Bai, H., Yu, A., *et al.*: 'Comparative study of compressive classification-based approaches for face recognition'. Proc. IEEE Conf. on Industrial Electronics and Applications (ICIEA), Hefei, China, 2016

[3] Turk, M., Pentland, A.: 'Eigenfaces for recognition', *J. Cogn. Neurosci.*, 1991, **3**, (1), pp. 71–86

[4] Belhumeur, P.N., Hespanha, J.P., Kriegman, D.J.: 'Eigenfaces vs. Fisherfaces: recognition using class specific linear projection', *IEEE Trans. Pattern Anal. Mach. Intell.*, 1997, **19**, (7), pp. 711–720

[5] He, X., Niyogi, P.: 'Locality preserving projections'. Proc. Neural Information Processing Systems (NIPS), Vancouver, British Columbia, Canada, vol. 16, 2003

[6] He, X., Yan, S., Hu, Y.*, et al.*: 'Face recognition using laplacianfaces', *IEEE Trans. Pattern Anal. Mach. Intell.*, 2005, **27**, (3), pp. 328–340

[7] Shen, L., Bai, L.: 'A review on Gabor wavelets for face recognition', *Pattern Anal. Appl.*, 2006, **9**, (2–3), pp. 273–292

[8] Wright, J., Yang, A.Y., Ganesh, A.*, et al.*: 'Robust face recognition via sparse representation', *IEEE Trans. Pattern Anal. Mach. Intell.*, 2009, **31**, (2), pp. 210–227

[9] Candès, E.J.: 'Compressive sampling'. Proc Int. Congress of Mathematicians, Madrid, Spain, vol. 3, 2006

[10] Donoho, D.L.: 'Compressed sensing', *IEEE Trans. Inf. Theory*, 2006, **52**, (4), pp. 1289–1306

[11] Deng, W., Hu, J., Guo, J.: 'In defense of sparsity based face recognition'. Proc. IEEE Conf. on Computer Vision and Pattern Recognition (CVPR), Portland, Oregon, USA, 2013

[12] Gao, Y., Ma, J., Yuille, A.L.: 'Semi-supervised sparse representation based classification for face recognition with insufficient labeled samples', *IEEE Trans. Image Process.*, 2017, **26**, (5), pp. 2545–2560

[13] Donoho, D., Tsaig, Y.: 'Fast solution of ℓ_1 -norm minimization problems when the solution may be sparse', *IEEE Trans. Inf. Theory*, 2008, **54**, (11), pp. 4789–4812

[14] Tropp, J.A., Gilbert, A.C.: 'Signal recovery from random measurements via orthogonal matching pursuit', *IEEE Trans. Inf. Theory*, 2007, **53**, (12), pp. 4655–4666

[15] Zheng, Z., Yang, F., Tan, W.*, et al.*: 'Gabor feature-based face recognition using supervised locality preserving projection', *Signal Process.*, 2007, **87**, (10), pp. 2473–2483

[16] Abdel-Sayed, M.M., Khattab, A., Abu-Elyazeed, M.F.: 'Fast matching pursuit for wideband spectrum sensing in cognitive radio networks', *Wirel. Netw.* 2017

[17] Kartha, A.V., Bharadi, V.A.: 'Face recognition using orthogonal transform coefficients of hyperspectral face images'. Proc. IEEE Int. Conf. on Information Processing (ICIP), Québec City, Canada, 2015

[18] Forczmański, P., Kukharev, G., Shchegoleva, N.: 'Simple and robust facial portraits recognition under variable lighting conditions based on two-dimensional orthogonal transformations'. Image Analysis and Processing – ICIAP, Naples, Italy, 2013 (LNCS, **8156**)

[19] Hassaballah, M., Aly, S.: 'Face recognition: challenges, achievements and future directions', *IET Comput. Vis.*, 2015, **9**, (4), pp. 614–626

[20] Daugman, J.G.: 'Uncertainty relation for resolution in space, spatial frequency, and orientation optimized by two-dimensional visual cortical filters', *JOSA A*, 1985, **2**, (7), pp. 1160–1169

[21] Chung, K.-C., Kee, S.C., Kim, S.R.: 'Face recognition using principal component analysis of Gabor filter responses'. Proc. IEEE Int. Workshop on Recognition, Analysis, and Tracking of Faces and Gestures in Real-Time Systems, Corfu, Greece, 1999

[22] Liu, C., Wechsler, H.: 'Gabor feature based classification using the enhanced fisher linear discriminant model for face recognition', *IEEE Trans. Image Process.*, 2002, **11**, (4), pp. 467–476

[23] Fernandes, S.L., Bala, G.J.: 'Recognizing facial images using ICA, LPP, MACE Gabor filters, score level fusion techniques'. Proc. Int. Conf. on Electronics and Communication Systems (ICECS), Coimbatore, India, 2014

[24] Boyd, S., Vandenberghe, L.: '*Convex optimization*' (Cambridge University Press, London, U.K., 2004)

[25] Vo, N., Vo, D., Challa, S.*, et al.*: 'Compressed sensing for face recognition'. Proc. IEEE Symp. on Computational Intelligence for Image Processing, Nashville, Tennessee, USA, 2009

[26] Yang, J., Luo, L., Qian, J.*, et al.*: 'Nuclear norm based matrix regression with applications to face recognition with occlusion and illumination changes', *IEEE Trans. Pattern Anal. Mach. Intell.*, 2017, **39**, (1), pp. 156–171

[27] Yang, W., Li, J., Zheng, H.*, et al.*: 'A nuclear norm based matrix regression based projections method for feature extraction', *IEEE Access*, 2018, **6**, pp. 7445–7451

[28] Lin, J., Chiu, C.-T.: 'Low-complexity face recognition using contour-based binary descriptor', *IET Image Process.*, 2017, **11**, (12), pp. 1179–1187

[29] Gabor, D.: 'Theory of communication. Part 1: the analysis of information', *J. Inst. Electr. Eng. Part III: Radio Commun. Eng.*, 1946, **93**, (26), pp. 429–441

[30] Roweis, S.T., Saul, L.K.: 'Nonlinear dimensionality reduction by locally linear embedding', *Science*, 2000, **290**, (5500), pp. 2323–2326

[31] Chang, Y., Hu, C., Turk, M.: 'Manifold of facial expression'. Proc. IEEE Int. Workshop on Analysis and Modeling of Faces and Gestures (AMFG), 2003, pp. 28–35

[32] Belkin, M., Niyogi, P.: 'Laplacian eigenmaps and spectral techniques for embedding and clustering'. Proc. Neural Information Processing Systems (NIPS), Vancouver, British Columbia, Canada, vol. 14, 2001, pp. 585–591

[33] Eldar, Y.C., Kutyniok, G.: '*Compressed sensing: theory and applications*' (Cambridge University Press, London, U.K., 2012)

[34] Abdel-Sayed, M.M., Khattab, A., Abu-Elyazeed, M.F.: 'Adaptive reduced-set matching pursuit for compressed sensing recovery'. Proc. IEEE Int. Conf. on Image Processing (ICIP), Phoenix, Arizona, USA, 2016

[35] Abdel-Sayed, M.M., Khattab, A., Abu-Elyazeed, M.F.: 'RMP: reduced-set matching pursuit approach for efficient compressed sensing signal reconstruction', *J. Adv. Res.*, 2016, **7**, (6), pp. 851–861

[36] Björck, A.: '*Numerical methods for least squares problems*' (SIAM, USA, 1996)

[37] Martínez, A.M.: 'The AR face database'. CVC Technical Report 24, 1998

[38] Martínez, A.M., Kak, A.C.: 'PCA versus LDA', *IEEE Trans. Pattern Anal. Mach. Intell.*, 2001, **23**, (2), pp. 228–233

[39] Georghiades, A.S., Belhumeur, P.N., Kriegman, D.J.: 'From few to many: illumination cone models for face recognition under variable lighting and pose', *IEEE Trans. Pattern Anal. Mach. Intell.*, 2001, **23**, (6), pp. 643–660

[40] Yang, A.Y., Sastry, S.S., Ganesh, A.*, et al.*: 'Fast ℓ_1 -minimization algorithms and an application in robust face recognition: a review', IEEE Int. Conf. on Image Processing (ICIP), Hong Kong, 2010