

Face Recognition Using a New Compressive Sensing-based Feature Extraction Method

Mehdi Banitalebi-Dehkordi · Amin Banitalebi-Dehkordi · Jamshid Abouei · Konstantinos N. Plataniotis

Received: date / Accepted: date

Abstract This paper proposes a novel face recognition algorithm that utilizes a sparse Fast Fourier Transform (FFT)-based feature extraction method. In our algorithm, we use Compressive Sampling (CS) theory two times. First, in the feature extraction process for extracting the feature vectors from a face images, and second, in the classification process where the CS reconstruction is used for selecting true classes. As a result, a significant reduction in the dimensionality of the signals is achieved. Extensive and comparative experiments have been conducted to evaluate the performance of the proposed scheme. The experiment results show that the combined Compressive Sensing and Sparse Representation Classification (SRC) achieves a high recognition accuracy, while maintaining a reasonable computational complexity.

Keywords Compressive sensing · feature extraction · neural networks · SRC classifier · FFT

M. Banitalebi-Dehkordi
Yazd University, Yazd, Iran

A. Banitalebi-Dehkordi
University of British Columbia, Vancouver, BC, Canada
E-mail: dehkordi@ece.ubc.ca

J. Abouei
Yazd University, Yazd, Iran

K. N. Plataniotis
University of Toronto, Toronto, Canada

1 Introduction

Face Recognition (FR) has attracted an enormous amount of attention in recent years, as it has a wide variety of applications in identification or verification of facial images. Since, FR algorithms are accurate and harmless, they are being used in applications such as access control, smart cards, information security, and law enforcement [23]·[13]·[15]. With the advancements in other related technologies, even more doors have been opened for advanced face recognition systems. For example, Reference [49] proposes a robust method of labeling mobile images on cloud platforms. As another example, reference [50] proposes an efficient person re-identification algorithm, in which pedestrians are identified across multiple cameras with different viewing geometry.

While there exists a huge body of work dealing with designing reliable FR systems, these systems are still facing some difficulties, e.g., when human faces look different in various lighting conditions or facial expressions. The majority of the existing literature employs shallow or deep learning of face recognition. Traditional algorithms based on feature extraction utilize local, or non-local facial features to train a model that will later be used to recognize an unseen face image [21]·[40]. Deep learning approaches on the other hand do not generate features, but instead they use the entire face image database to learn and predict unseen face images [51]·[52]·[53]·[54]·[55]·[56]. State-of-the-art deep learning methods have reported to achieve very high recognition accuracy [51].

Moreover, most of the current FR approaches assume that multiple images of a same face are available for training their algorithms [37]·[33]. However, in practical situations, there are only limited number of images available for the same face. Moreover, employing more images would increase the complexity and the processing time of the algorithm. In the case where only one sample image is available for training of the system, it is difficult to recognize the person's face in different lighting conditions. Reference [32] demonstrates that the performance of FR algorithms degrades even more for the "one sample per person" case when there is no limitation on the lighting conditions of the images under test. To solve this problem, some researches have focused on employing local features in FR [37]·[22]·[30]·[25]. Authors in [31] propose a face description and similarity measuring technique for the Visual Similarity Retrieval (VSR) in single model face databases. Instead of using the local operations on isolated points, the algorithm in [31] employs the Directional Corner Point (DCP) matching in which directional information showing the connectivity to neighbors is utilized for the similar point. Unlike neural network approaches that require multiple model faces per person to train the system for optimal setting [1], the DCP method is suitable for the single model face database retrieval as it is fast in computation. For graph matching techniques in [11]·[12]·[16], feature vectors are extracted at points on the face image which contains more information. In most feature based methods, facial features are assumed to be the eyes, nose and mouth. In [11]·[12]·[16], however, the locations and the number of feature points are not fixed, therefore, the number

of feature vectors and their locations can vary to represent different facial characteristics of different human faces. In this way, feature points are located around the main facial features (e.g., eyes, nose and mouth) as well as around the special facial features of an individual such as dimples.

Unlike the first group of researchers who are interested in local features, some other works in the literature indicate that instead of using only local features, the entire face should be taken into account. This line of work has been discussed in [21] where the authors employ the principal component analysis to enrich the information extracted from the facial images, or in [26], two-dimensional principal component analysis is used. For such scenarios, a large number of features are used for extracting the information which makes the computational cost of the system increase. To reduce the computational complexity of such a FR process, some sort of dimensionality reduction mechanisms are used before data classification. Linear Predictive Coding (LPC) [41] and Linear Discriminant Analysis (LDA) [36] are two classic approaches for such a dimensionality reduction. In addition to those classic methods, *Compressive Sensing* (CS) is an efficient mechanism used for decreasing the dimensionality of feature vectors. In comparison to other feature extraction methods, in a CS-based approach, the feature extraction is performed by extracting random samples. CS-based approaches generally result in lower-dimensional measurements of images, without significantly compromising the recognition performance.

In most of the previous works on the CS-based face recognition tasks, e.g., [43]-[44]-[40], face images are processed in the frequency domain and then, they are classified based on the gender of human subjects appearing in the image. In [44], different images of the same gender are considered as an ensemble of inter-correlated signals. These signals are assumed to be sparse due to variations in faces with respect to the whole image. The authors in [44] exploit this sparsity using CS which enables them to grossly represent images of a given gender by two kinds of features: one represents common features of different faces and another which denotes different faces in all training samples. The first and the second features are combined by a cascaded filter for robust gender recognition. Reference [40] also views the different images of the same subject as an ensemble of inter-correlated signals where the signals are assumed to be sparse on the variations in expressions [40]. The proposed scheme in [40] employs this sparsity using distributed CS theory to represent training images of a given subject by only two feature images: one that captures the holistic (or common) features of the face, and the other which captures different expressions in all training samples.

Some of the CS-based face recognition algorithms use Sparse Representation based Classifiers (SRCs) for classification [43], in which a query sample is encoded as a sparse linear combination of all training samples. Then, the face images are classified by evaluating the classification error, i.e., selecting the class with the minimum representation error. It is demonstrated in [43] that the ℓ_1 -norm sparsity constraint on encoding of the face matrix's coefficients plays a key role in the success of SRCs. By synthesizing the algorithm

presented in [43] with our Compressive Sensing based two-levels FFT (CS-2FFT) feature extraction process, we propose a robust and fast CS-based face recognition algorithm to generate a sparse feature vector for each image. In this paper, we propose a new CS-based technique that utilizes the FFT to extract a sparse feature vector which has the least possible number of non-zero transform coefficients. In the proposed method, a Hamming windowing function is applied to the image, and then the FFT coefficients resulted by taking the FFT from each window are added together. Therefore, we will end up having one number for each window corresponding to the summation of the normalized FFT coefficients in that window and this gives us a vector for each image. We take the second FFT from this vector and show that if we take a random sample of the resulting vector, it would contain enough information to discriminate the corresponding facial image from other images in a database.

With the above insights in mind, this paper presents for the first time a new modeling and analysis framework for the face recognition task and a sparse FFT-based feature extraction mechanism for characterization of face images. We represent the image as a vector obtained by connecting the image columns sequentially together. To further reduce the dimensionality of our face recognition problem, a CS-2FFT feature vector is synthesized for each image vector. This feature extraction process takes into account both locally extracted features of the image such as those mentioned in the VSR [31] and F-FR [11], and globally extracted features such as the ones in FLDA [27], CLDA [24], and FFR [18]. The proposed classifier is based on the principle of the SRC with an additional random measurement on the extracted features to reduce the number of dimensions in the image vector which is fed to the classifier. In the proposed face recognition algorithm, we use the CS theory two times. First, in the feature extraction process for extracting the feature vectors from the face images (described in Section 3). This process does not deploy the CS reconstruction, thus, there is no need for the information of the used dictionaries. Second, as it will be mentioned in Section 4, in our classification process, we use the CS reconstruction for selecting the true class. As a result, a significant reduction in the dimensionality of the image vectors is achieved. Extensive experiments have been conducted to evaluate the performance of the proposed scheme in comparison to the existing shallow learning based methods such as the robust VSR in single model face databases [31] and the FLDA applicable to the face recognition with one sample per person [27]. Experimental results show that Compressive Sensing combined with the Sparse Representation Classification (SRC) [43] achieve a high recognition accuracy on Olivetti Research Laboratory (ORL) face database (96%), Labeled Faces in the Wild (LFW) home face database (42% - when trained/tested only on a subset of the database), Extended Yale B face database (95.9%), and Alex Martinez and Robert Benavente face database (AR) (95.6%), while maintaining a reasonable computational complexity.

The rest of this paper is organized as follows: Section 2 reviews the compressive sampling strategy, Section 3 explains the proposed sparse feature extraction mechanism, Section 4 elaborates on the sparse representation-based

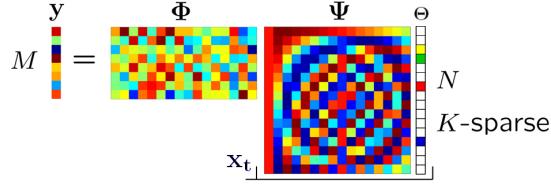


Fig. 1: Compressive Sensing in matrix notation.

classification, Section 5 provides performance evaluations, experimental results and interpretation of the findings, and finally Section 6 concludes the paper.

2 Overview of Compressive Sensing Strategy

In our proposed approach, Compressive Sensing is being used as an efficient mechanism for dimensionality reduction of the feature vectors. Thus, it is essential to bring a brief overview of the utilized Compressive Sensing scheme in this section. Illustrated in Fig. 1, suppose \mathbf{x}_t is a sparse vector of order K on the basis Ψ , i.e., \mathbf{x}_t has only K non-zero coefficients on basis matrix Ψ . In addition, suppose Φ denotes a $M \times N$ sampling matrix with $M \ll N$, whose rows are uncorrelated with columns of Ψ . If Θ is the sparse samples vector, then we define

$$\mathbf{x}_t = \Psi\Theta. \quad (1)$$

According to the Compressive Sensing theory, \mathbf{x}_t can be recovered using $M = O(K \log(N))$ non-adaptive samples as follows [38],[39]:

$$\mathbf{Y} = \Phi\mathbf{x}_t = \Phi\Psi\Theta = \mathbf{D}\Theta, \quad (2)$$

where \mathbf{Y} is an $M \times 1$ vector representing the samples and \mathbf{D} is an $M \times N$ matrix. To recover the signal, the sparse coefficients vector Θ is computed using the following ℓ_0 -norm optimization problem:

$$\min \|\Theta\|_0 \quad s.t. \quad \mathbf{Y} = \Phi\mathbf{x}_t = \mathbf{D}\Theta. \quad (3)$$

Solving (3) in its general form is complicated, thus, ℓ_1 -norm is practically used instead of the ℓ_0 -norm as follows [11],[12]:

$$\min \|\Theta\|_1 \quad s.t. \quad \mathbf{Y} = \Phi\mathbf{x}_t = \mathbf{D}\Theta. \quad (4)$$

Therefore, instead of taking N samples from a signal which is sparse of order K , only M perpendicular samples of the signal are taken. This means that less number of samples are used in the entire face recognition process. Independent identically distributed (*i.i.d*) random Gaussian vectors (0, 1) and Bernoulli vectors (± 1) are among popular random vectors which are perpendicular within each other and therefore these vectors can be utilized for sampling of the image.

3 Proposed Feature Extraction Algorithm

A schematic process of the proposed face recognition approach is illustrated in Fig. 2. In this figure, we employ two main steps for the recognition task. In the first step, known as the learning process, we extract CS-2FFT feature vectors from each class symbol, while in the second step, we extract CS-2FFT feature vectors from new test pictures and calculate the minimum error between test feature vectors and class symbol feature vectors. For the above FR task, in this section, we introduce the CS-2FFT feature extraction process employed in both steps.

To extract a sparse feature vector from t^{th} face image, we first create a $M \times 1$ vector \mathbf{x}_t by connecting the image columns sequentially together. This vector is then partitioned using 50 percent overlapped Hamming windows meaning that the first half of each window is overlapped with the previous window, while the second half is overlapped with the next window. Applying the FFT to i^{th} window and normalizing the resulted coefficients by their maximum values so that the normalized magnitudes are between 0 and 1, we make a $M \times 1$ vector $\mathbf{x}_{nor_i}^{(t)}$. Since, t^{th} face vector \mathbf{x}_t is similar to a summation of sinusoidal waves, we can use the \mathbf{x}_t 's FFT coefficients or equivalently the spectrogram as a starting point in our CS-2FFT feature extraction process to extract $\mathbf{x}_{F(w_i)}^{(t)}$. It should be noted from $\mathbf{x}_{F(w_i)}^{(t)}$ that the size of the used FFT is equal to the length of \mathbf{x}_t . In fact, to make better time invariant properties in the output of the “summation of window coefficient” block in Fig. 2, local features, i.e., the FFT coefficients of each window, are added together. In the next step, we add up these transform coefficients for each window to assign one value per window which represents the small-time feature of that window. We put the associated window values together to make a $B \times 1$ vector $\mathbf{x}_{sum_i}^{(t)}$ and take another FFT from the resulted signal. The output of this step represents the long-time feature of the original image.

Note that the Fourier analysis is a process that decomposes a face image vector into the summation of a series of sine and cosine waves with different amplitudes, frequencies and phases. Each Fourier transform coefficient can be interpreted in the frequency domain corresponding to the original signal in the spatial domain. The first FFT operates as a moving window filter that is convolved with the image. Window summation aggregates the filtered window samples as a part of the convolution. This provides a notion of local changes (50% overlapping windows) of texture variations for each window. Then the second FFT extracts the longer term non-local texture variations, and results in a sparse representation. Another method of extracting texture and orientation of an image through spectral exploitation, that is widely used in the literature, is to use Gabor filters. However, the 2FFT approach results in a sparse representation that fits better in the rest of the proposed solution.

It is also worth noting that due to the two FFT stages, local and non-local variations are both taken into account. As a result, theoretically it shouldn't make a noticeable difference to construct a vector signal from image rows

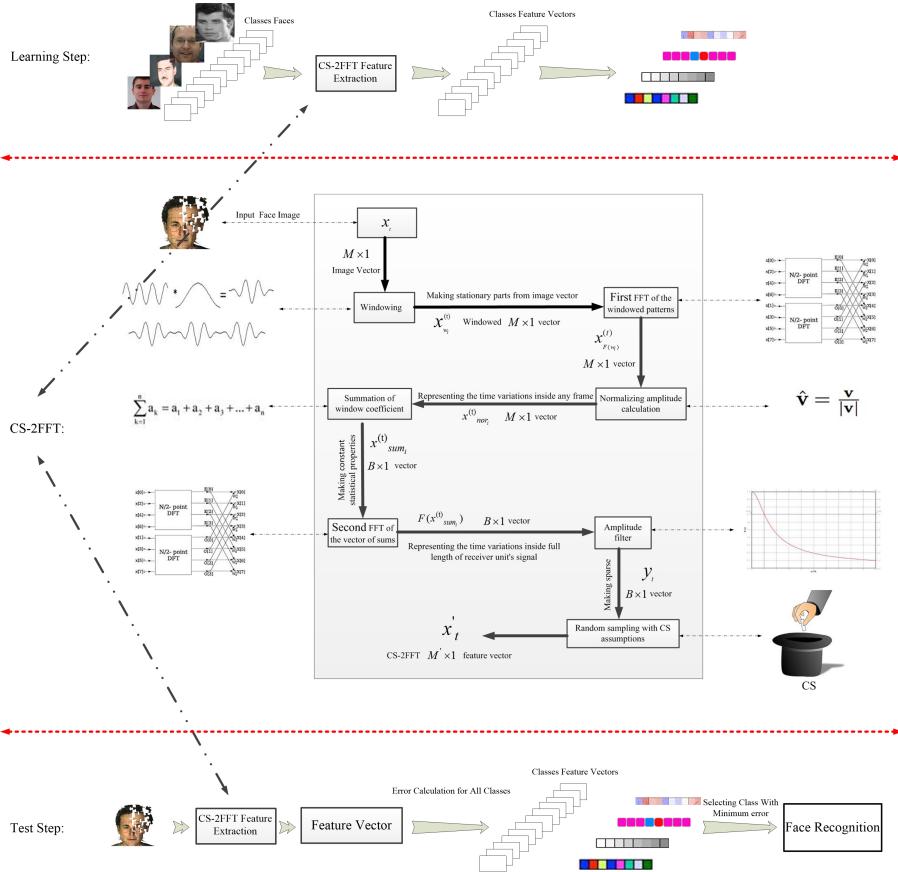


Fig. 2: Block diagram of the proposed face recognition scheme.

instead of image columns. This was also verified statistically through our experiments. However, it should be disclosed that no other way of windowing was tested (e.g. zigzag scan, 2D block windows, or etc.). The main motivation behind our choice was to achieve a very fast parallelizable implementation, as we do not expect to see a noticeable difference in the sparse representation due to the 2FFT approach.

Deploying the second FFT on the vector of feature samples of the time-invariant data results in the $B \times 1$ vector $F(\mathbf{x}_{sum_i}^{(t)})$ representing the obtained feature vector. Similar to the first FFT, it is sensible from $F(\mathbf{x}_{sum_i}^{(t)})$ that the size of the second FFT is equal to the length of $\mathbf{x}_{sum_i}^{(t)}$. It is worth mentioning that the second FFT provides a sparse representation of the feature vector obtained by applying a predetermined amplitude threshold. For this case, the length of this feature vector is equal to the number of windows. Note that while the first FFT represents the time variations inside any frame of the face vector, i.e., local properties of the face image, the second FFT is deployed to indicate the variations of the local properties during consecutive frames. To have an image vector that is as sparse as possible, we feed $F(\mathbf{x}_{sum_i}^{(t)})$ into an amplitude filter which sets the values that are less than a threshold to zero.

This amplitude threshold is used to clip the values of the normalized feature elements. As we will show in Section 5, a proper choice of the threshold value results in a higher degree of sparseness in the feature vector that yields a shorter CS-2FFT vector.

In the last step, features are extracted randomly by sampling from the sparse vector \mathbf{y}_t . The sampling process is performed by multiplying the sparse $B \times 1$ vector \mathbf{y}_t by a random Gaussian sampling matrix which leads to the $M' \times 1$ vector \mathbf{x}_t . Number of the samples depends on the sparsity level of the \mathbf{y}_t vector. We empirically choose this parameter in our numerical evaluations in Section 5.

4 Compressive Sensing based Classifier

Compressive Sensing based classifiers utilize properties of the sparse signals to develop classification algorithms. In this section, we provide more insight through a classification method that take the advantage of this sparsity in our face recognition algorithm (see Fig. 2).

Suppose $\mathbf{v}_{i,j} \in \mathbb{R}^M$ is the j^{th} extracted sample vector from the face images in the i^{th} class. Also, we assume that there are enough sample vectors in the i^{th} class in the form of $[\mathbf{v}_{i,1}, \mathbf{v}_{i,2}, \dots, \mathbf{v}_{i,n_i}]$, where n_i represents the number of samples in i^{th} class. If $\mathbf{x}^{(t)} \in \mathbb{R}^M$ is t^{th} sample vector extracted from a test image in the same class as $\mathbf{v}_{i,j}$, then it is approximately represented by the following linear expression:

$$\mathbf{x}^{(t)} = \sum_{j=1}^{n_i} \Theta_{i,j} \mathbf{v}_{i,j}, \quad (5)$$

where $\Theta_{i,j}$'s denote the weighting parameters. Let's define \mathbf{A} as a matrix whose column \mathbf{A}_i is filled with the feature vector corresponding to the i^{th} class $\mathbf{v}_{i,j}$, then \mathbf{A} would be a function similar to Ψ in (1) and (2). This can be written as

$$\mathbf{x}^{(t)} = \mathbf{A}\Theta \in \mathbb{R}^M, \quad (6)$$

where Θ is an expanded feature vector defined as $\Theta = [\Theta_{i,1}, \Theta_{i,2}, \dots, \Theta_{i,n}] \in \mathbb{R}^n$ and $n = \sum_{i=1}^K n_i$, where K is the number of classes. More precisely, the expanded feature vector is a coefficient vector whose elements are zero except for the elements associated with the i^{th} class [16]. To find the sparse representation vector $\widehat{\Theta}$ for any input vector \mathbf{x}_t , we use the optimization problem in (3) or (4). In the ideal case, all non-zero elements of $\widehat{\Theta}$ correspond to the columns of \mathbf{A} that are associated with the i^{th} class. However, due to the existence of the noise and the dissimilarity between the input signal and each class signals, the ideal case would not occur. To resolve this problem, one feature function $\delta_i : \mathbb{R}^n \rightarrow \mathbb{R}^n$ is defined for the i^{th} class such that $\delta_i(\widehat{\Theta}) \in \mathbb{R}^n$ is a new vector whose non-zero elements are the corresponding elements of $\widehat{\Theta}$ for the i^{th} class.

Assuming that the test data belongs to the i^{th} class, Θ can be estimated from (6) as follows:

$$\hat{\mathbf{x}}_i^{(t)} = \mathbf{A}\delta_i(\hat{\Theta}). \quad (7)$$

Once $\hat{\Theta}_i$'s and $\hat{\mathbf{x}}_i^{(t)}$'s are evaluated for $i = 1, \dots, K$, then \mathbf{x}_t falls within a class with the least estimation error, i.e.,

$$\min r_i(\mathbf{x}_t) = \|\mathbf{x}_t - \hat{\mathbf{x}}_i^{(t)}\|_2, \quad (8)$$

where $r_i(\mathbf{x}_t)$ is the least estimation error of i^{th} class.

Algorithm 1 summarizes our proposed face recognition scheme. In the next section, we provide some experiment results to evaluate the proposed algorithm in Algorithm 1.

Algorithm 1 Proposed CS-2FFT-based face recognition algorithm

```

1: procedure
2:   Input: Face image  $x_t$  and face database
3:   Output: Face recognition
4:   Algorithm:
5:   for  $k = 1 : K$  do            $\triangleright$  K determines the number of face classes
6:     Extract features from K classes for the training data
7:     Feature extraction:
8:       Construct overlapping Hamming windows
9:       Apply the first FFT
10:      Normalize and add up coefficients to get local features
11:      Apply the second FFT to produce the non-local features
12:      Apply the amplitude filter
13:      Create a sparse matrix  $\mathbf{A}$  using the extracted features
14:      Take a vector sample of  $\mathbf{A}$  using the Compressive Sensing
         method through multiplying  $\mathbf{A}$  by a sampling matrix  $\Phi$ 
         (e.g., a random Gaussian)
15:      Normalize the sampled matrix  $\mathbf{x}_{F(w_i)}^{(t)}$  so that its  $\ell_2$ -norm equals
         to unity. Then, the signal recovery is performed for all the classes
16:      Estimation error is evaluated for all the classes:  $r_i(\mathbf{x}_t) = \|\mathbf{x}_t - \hat{\mathbf{x}}_i^{(t)}\|_2$ 
17:      Best class is the one with the least amount of error:  $\min r_i(\mathbf{x}_t)$ 

```

5 Experimental Results and Discussion

In this section, extensive and comparative experiments have been conducted to evaluate the performance of the proposed scheme when compared to the state-of-the-art methods such as the robust VSR in gray-scale and single model face databases [31] and the FLDA applicable to the face recognition [27]. All face recognition experiments are performed over ORL [47], LFW [34], Extended Yale B face database, and Alex Martinez and Robert Benavente face database (AR). The algorithms are tested using MATLAB version 8.1 on Windows 8 platform and a Pentium 4 machine with 4 GB of memory. The following describes the details of the experimental setup and the results.



Fig. 3: Snapshots of the ORL dataset.



Fig. 4: Snapshots of the LFW dataset.

5.1 Face Datasets

The ORL dataset consists of 400 face images of 40 people. All images of each person have been taken in different times, poses, and lighting situations. Also, images of each person are with glasses, closed/open eyes, or with/without smile. Fig. 3 shows snapshots of the ORL dataset including the gray level images with 8 bits brightness intensity depth and the resolution of 92×112 . For all face images in the ORL database, we randomly select five image for each person as the training sample and the remaining images are considered as the test images set.

The images in the LFW dataset are selected from news articles, and include all of the parameter variations mentioned above. The LFW dataset in our experiments contains more than 10,000 face images collected from the web, and 1,500 subjects which have two or more different face photos in this dataset. These images have been arranged into resolution of 64×64 with a commercial face alignment software [9]. Hence, our experiments on this dataset are considered as fully automatic face recognition. Fig. 4 shows 16 sample images of five subjects from the LFW database. It is seen from this figure that LFW contains many natural variations of poses, expressions, lighting, and misalignments. Therefore, face recognition for this dataset is more challenging than the images in the ORL database which have been collected under controlled conditions. In our experiments, only the first ten images per each subject are chosen. Moreover, for this dataset, the first image of each person is selected as the class symbol and the remaining images are considered as the class members. In addition, each class has ten members, where there are 151 members

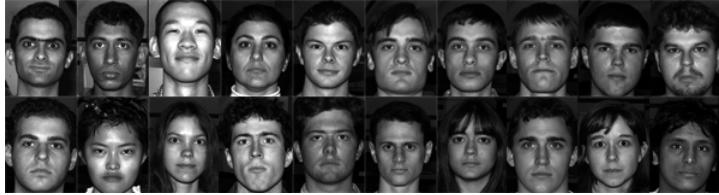


Fig. 5: Snapshots of the Extended Yale B dataset.



Fig. 6: Snapshots of the AR dataset.

in total. Among ten images of each person, five of them are used for training and the rest are used for the test. Training images are chosen such that they have maximum uncorrelated information. Therefore, 200 images are used for the training from which 200 feature vectors are extracted.

The Extended Yale B database contains 2,414 images of 38 subjects (around 63 images per person). The cropped and normalized 192×168 face images have been captured under various laboratory-controlled lighting conditions [3]. Similar to the other utilized databases, for each class we use half of the images for training (i.e., about 30 images per subject), and the other half is used for testing. Randomly choosing the training set ensures that our results will not depend on any special choice of the training data. Fig. 5 shows sample images of different subjects in the Extended Yale B database.

The AR database consists of over 4,000 images of 126 subjects. For each subject, 26 pictures are selected as training set and another 26 images are used for testing set [2]. Compared to the Extended Yale B database, the images in the AR database contain more parameter variations, e.g., a wider range of illumination change, expressions, and facial disguises. For each subject, 14 images with illumination and expression changes are selected; seven images for training, and the other seven for testing. The images are cropped to the dimension of 165×120 and are converted to grayscale. Fig. 6 shows sample images of different subjects in the AR database. This database is substantially

more challenging than the Extended Yale B database since the number of subjects is now increased to 126, however, the training images for the AR dataset is reduced to only seven images per subject.

5.2 Experimental Setup

The experiments consist of two sets. The first benchmark compares the accuracy and the speed of the algorithms for performing the face detection. For this case, the sparse source vector and the under determined linear system are generated based on image columns. The second benchmark measures the face recognition accuracy in noisy situation where the training images are taken from AR, Extended Yale B, LFW and ORL face recognition databases.

5.3 Length of The Extracted Feature Vector

The length of feature vectors in all classical face recognition algorithms directly affects the computational complexity and the performance of the methods. As explained in Section III, in our proposed feature extraction process, we can control the length of the feature vector by adjusting the amplitude threshold. This amplitude threshold is used to clip the values of the normalized feature elements. Fig. 7 shows the length of our proposed feature vector in terms of the amplitude threshold for ORL, LFW, AR, and Extended Yale B databases. It is shown in Fig. 7 that a proper choice of the threshold value results in a higher degree of sparseness in the feature vector that yields a shorter CS-2FFT vector.

Moreover, in order to observe the effect of the feature length in the face recognition accuracy, we illustrate the face recognition percentage and the feature dimension in terms of the amplitude threshold in Fig. 8. As explained in Subsection 5.1, the AR database is more challenging than other used databases, therefore, we utilize the AR database in this part of simulations. The results in both figures 7 and 8 show that higher values of the amplitude threshold yields a shorter length of the feature vector, while the face recognition accuracy is not significantly affected. To achieve the best trade-off between the length of the feature vector and the face recognition accuracy, in our proposed face recognition algorithm, we empirically choose a point from this trade-off to 0.45.

5.4 Experimental Results and Interpretations

In our simulation setup, we use the False Recognition Rate (FRR) and the False Acceptance Rate (FAR) metrics. The FRR is a measure of the likelihood that the biometric security system will incorrectly accept an access attempt by an unauthorized user. More precisely, the FAR of a system is stated as the ratio of the number of false acceptances divided by the number of identification

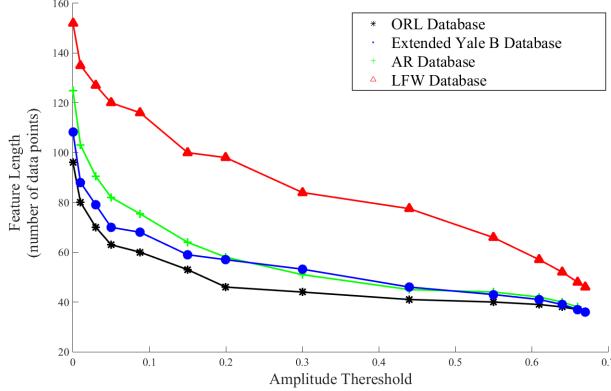


Fig. 7: Length of the proposed feature vector in terms of the amplitude threshold for various datasets.

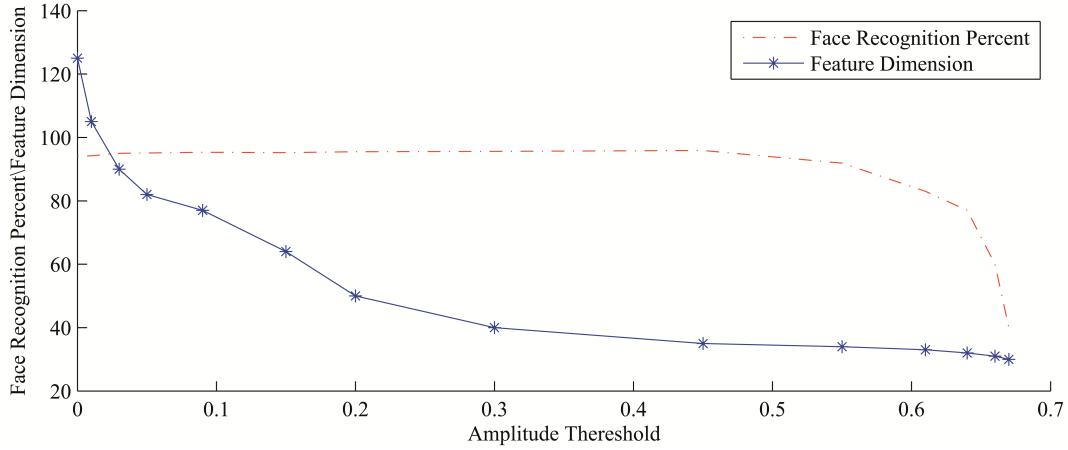


Fig. 8: Performance of the proposed face recognition algorithm in terms of the amplitude threshold for the AR database and SNR= 5 dB.

attempts. This metric is typically obtained as the ratio of the number of false rejections divided by the number of identification attempts. For our databases, the results show $FRR = 3.0\%$ and $FAR = 1.4\%$. In addition, the correct classification rate is 95, 6% which only gives 4, 4% misclassification. This is an interesting result, since the images of each person in the AR database are in a wide range of conditions and noticeably different.

Tables 1 and 2 compare the accuracy and the complexity $e(t)$ of the proposed face recognition algorithm with some classical recognition schemes such as Powel-Beale CGA [42], scaled CGA [29], Powel-Beale CGA and PCA [35],

Table 1: Face recognition rate in different methods for SNR= 5 dB.

Algorithm	ORL dataset	LFW dataset	Extended Yale B dataset	AR dataset
PCA	86.2%	18.5%	84.6%	83.9%
LDA	87.4%	19.1%	86.2%	85.1%
Scaled CGA	90%	20.4%	89.2%	88.6%
F-CS	92%	29.5%	91.3%	90.9%
Powel-Beale CGA	92%	30.4%	91.5%	91%
CA-SRC	93%	30.6%	92%	91.3%
EI-CS	93%	34.9%	92.7%	92.1%
Powel-Beale- CGA-PCA	94%	36.8%	93.2%	92.9%
Proposed method	96%	41.6%	95.9%	95.6%

CA-SRC [43], F-CS [44], PCA [8], LDA [7] and EI-CS [40]. To obtain the accuracy of the aforementioned algorithms in Table 2, we use the face recognition error rate. Note that the only true measure of the complexity of an algorithm is the mathematical measurement of the orders of operations it needs to complete. However, since almost all of the face recognition algorithms are so complex that makes such measurements impractical, we assess the algorithms complexity in terms of runtime. Table 2 shows the complexity as the average runtime of a recognition process for different face recognition algorithms in a local PC box (specifics above). It is also worth noting that some classical algorithms implementations such as PCA and LDA are available online [57] while others were implemented in-house as the implementation of such methods were not provided by their authors. We tried our best to ensure our in-house implementations match the original papers. It should be disclosed however that some detailed parameters and configurations could have been tuned slightly differently than what the original authors have used to generate their plots and performance reports.

As observed from these tables, our proposed face recognition algorithm has better accuracy along with a less complexity than other schemes for LFW, ORL, AR and Extended Yale B databases. These results come from the CS-2FFT feature extraction algorithm employed in the face recognition process, while e.g., in Powel-Beale CGA, CGA and PCA, the authors use other classical feature extraction algorithms for identifying faces. The accuracy of the proposed method also shows the robustness of our algorithm against the noise and different pose and lighting conditions [45]. In [45], the authors study the robustness of the feature extraction process by extracting the CS-2FFT from a simulated noisy sine wave and compare that with the feature vector extracted from the clear wave. Since other classical face recognition algorithms process a large amount of information provided in an image database, these algorithms usually suffer from a high complexity. In the case of a high resolution image database, this effect becomes even worse, while in our proposed face recognition algorithm, we utilize Compressive Sensing in both feature extraction and classification processes to significantly decrease the computational cost.

Table 2: Complexity $e(t)$ of various face recognition schemes for SNR= 5 dB.

Algorithm	ORL dataset	LFW dataset	Extended Yale B dataset	AR dataset
PCA	31.2 ms	34.6 ms	32.9 ms	34.1 ms
LDA	25.4 ms	28.1 ms	26.3 ms	28.4 ms
Scaled CGA	12.1 ms	14.4 ms	13.5 ms	16.1 ms
Powel-Beale CGA	9.4 ms	11.9 ms	10.6 ms	12.9 ms
F-CS	8.5 ms	10.1 ms	9.1 ms	11.5 ms
CA-SRC	7.6 ms	9.3 ms	8.4 ms	9.7 ms
EI-CS	6.9 ms	8.3 ms	7.4 ms	8.2 ms
Powel-Beale- CGA-PCA	5.4 ms	6.2 ms	5.8 ms	6.1 ms
Proposed method	3.2 ms	4.5 ms	3.8 ms	4.4 ms

We also examined the effect of the image resolution on the face recognition rate for the aforementioned algorithms. To this end, we down-sampled the 92×112 ORL images to $92/L \times 112/L$ images by down-scaling their resolutions. Fig. 9 illustrates the correct face recognition rate versus the image resolution for SNR= 5 dB, and for the ORL and LFW datasets. In this figure, the horizontal axis represents different values of $92/L$ and the vertical axis represents the face recognition rate. The results in Fig. 9 show that an increase in the image's width leads to an increase in the face recognition rate in all algorithms, while in the proposed algorithm, the growth is small than that of the other schemes. It means that the proposed algorithm is robust against the variation of the input image size. It is seen from Fig. 9 that our face recognition algorithm can accurately recognize a person from his/her images with different resolutions.

In the next step, we evaluate the relevance of sparsity for rejecting invalid test images for the AR data base. We again simulate 50% levels of occlusion by replacing a randomly chosen block of each test image with an unrelated image. However, in this experiment, we include only half of the subjects in the training set. Thus, half of the subjects in the testing set are new to the algorithm. We test the performance of our algorithm to determine whether a given test subject is in the training database by sweeping the amplitude threshold through a range [0, 1]. Fig. 10 shows the Receiver Operator Characteristic (ROC) curves of this experiment. For comparison, we also consider the outlier rejection by thresholding the Euclidean distance between the CS-2FFT feature vector extracted from the test image and the CS-2FFT feature vector extracted from the nearest training images within the proposed feature space. It is seen from Fig. 10 that at 50% occlusion, our proposed method still outperforms other assumed face recognition algorithms.

To complete our numerical evaluations, we examine the effect of the noise power on the performance of the face recognition algorithms shown in Tables 3 and 4. For this simulation setup, we add the Additive White Gaussian Noise (AWGN) to face images and apply the face recognition algorithms. The results in Tables 3 and 4 show that the proposed algorithm provides a superior

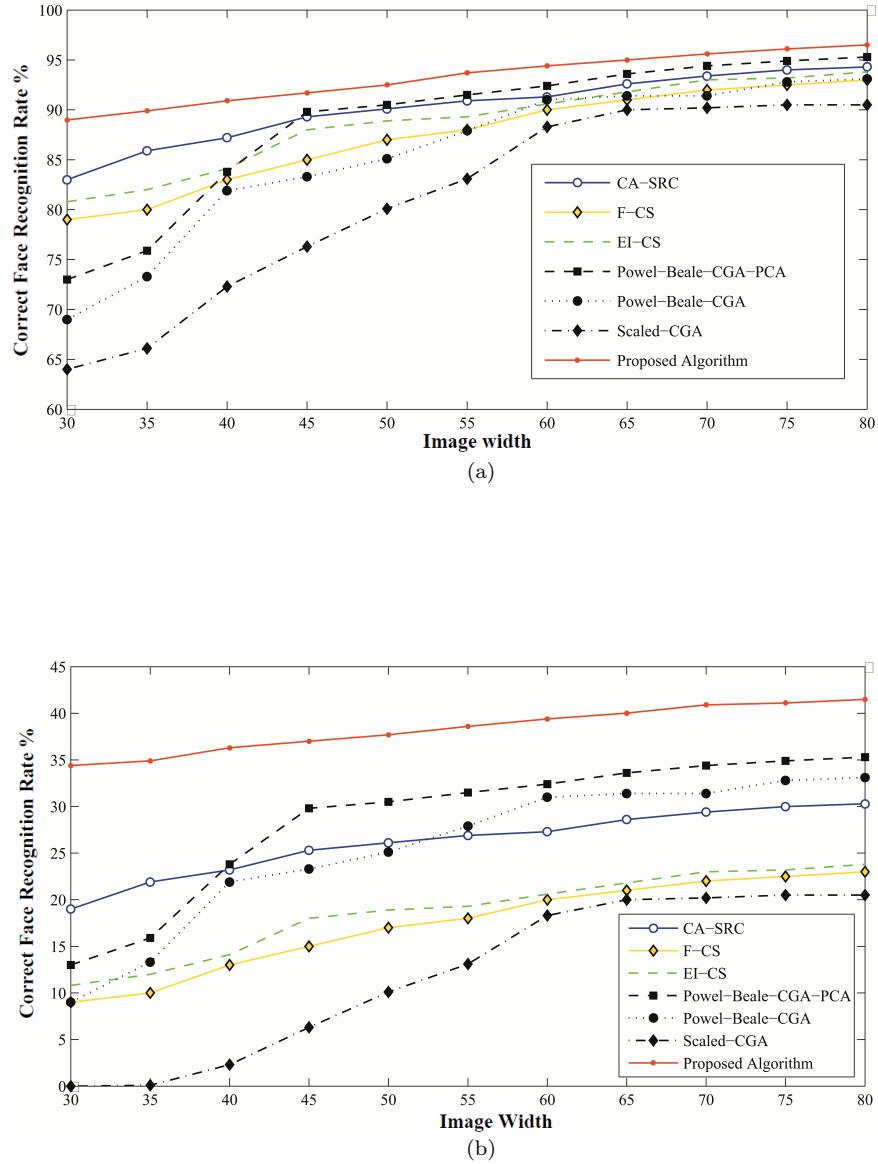


Fig. 9: Correct face recognition rate versus the image resolution for different algorithms with SNR= 5 dB and for a) ORL dataset, and b) LFW dataset.

accuracy and lower complexity than the other algorithms of Table 3 and 4, for different Signal to Noise Ratio (SNR) values (Note that our scheme is robust against noise variations).

Remark 1: Based on the experimental results on the used databases, it is concluded that *i*) for all the aforementioned datasets, the best performances of the proposed CS-2FFT based face recognition algorithm consistently exceed

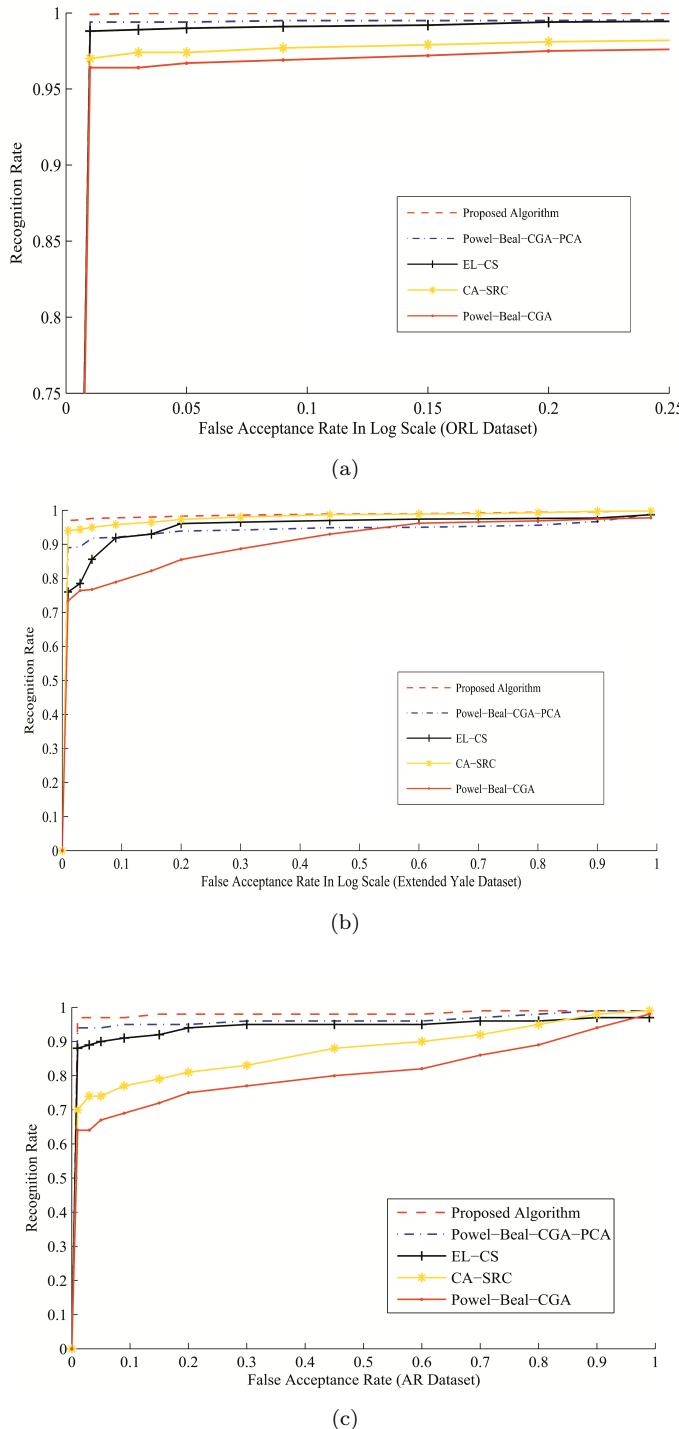


Fig. 10: ROC curves for outlier rejection (recognition rate in terms of false acceptance rate) for a) ORL dataset, b) Extended Yale B dataset, and c) AR dataset.

Table 3: Face recognition percentage in different SNRs and for ORL dataset.

SNR (dB)	Powel-Beal- CGA-PCA	Powel- Beal-CGA	Scaled- CGA	EL- CS	F- CS	CA- SRC	Ours
15 dB	94.6%	92.2%	90.2%	93.2%	92.2%	93.2%	97%
10 dB	94.6%	92.1%	90.2%	93.2%	92.1%	93.1%	96.7%
5 dB	94.5%	92.1%	90%	93.1%	92.1%	93.1%	96.4%
0 dB	94.5%	92%	90%	93.1%	92%	93%	96.2%
-5 dB	94%	92%	90%	93%	92%	93%	96%
-10 dB	93.5%	92%	89%	92%	91.2%	92.6%	95.9%
-15 dB	93%	91.8%	88%	91%	90.7%	92.1%	95.8%

Table 4: Face recognition percentage in different SNRs and for LFW dataset.

SNR (dB)	Powel-Beal- CGA-PCA	Powel- Beal-CGA	Scaled- CGA	EL- CS	F- CS	CA- SRC	Ours
15 dB	37.3%	31.4%	21.5%	26.2%	31.5%	32.9%	42.2%
10 dB	37.3%	31.2%	21.2%	26.1%	31.4%	32.1%	42.1%
5 dB	37.3%	30.9%	20.9%	25.9%	31.2%	31.5%	42.1%
0 dB	37.1%	30.6%	20.6%	25.4%	30.4%	30.9%	41.8%
-5 dB	36.8%	30.4%	20.4%	24.9%	29.5%	30.6%	41.6%
-10 dB	36.4%	30.1%	20.1%	24.9%	28.7%	30.4%	41.6%
-15 dB	36.1%	29.9%	19.8%	24.8%	28%	29.4%	41.5%

the best performances of Powel-Beale CGA, PCA and other classical algorithms at each individual feature dimension, *ii)* the performances of the other assumed face recognition algorithms depends strongly on the noise, while, as illustrated in Table 3, the performance of our proposed face recognition algorithm does converge in different SNR regimes.

Remark 2: It is worth mentioning that the basic SRC method in [43] does not incorporate any dimensionality reduction, while, in our scheme, we use the dimensionality reduction in the classification in accordance to the sparse feature extraction scheme. Since our proposed Compressive Sensing-based feature extraction process does not contain redundancies, it is suitable for the fast and reliable classifications.

Remark 3: It is worth noting that deep learning based approaches (Convolutional Neural Networks(CNNs) specifically) have been reported to achieve high recognition accuracy [51]·[52]·[53]·[54]·[55]·[56]. Evaluating the performance of such deep learning based methods (over different levels of SNR and other conditions tested above) is left outside the scope of this paper and will be explored as a part of future works.

Remark 4: Please also note that the experimental results for the LFW database shown above are based on training and testing on only a subset of the database (for the sake of faster processing). It is therefore expected to see a lower accuracy compared to the case when all of the database is used (reported by [51] some deep learning based methods achieve 96 + % accuracy when the entire LFW database is used).

Future Work: Future work includes the incorporation of deep learning based algorithms in the experimental comparisons, to evaluate the performance of these methods in the presence of different amounts of noise (SNR levels), size of datasets, number of face images per person, lighting conditions, and etc. We will also utilize the entire LFW database in the performance evaluations. This will particularly be important when evaluating the deep learning based methods, as they may require large amounts of labeled face images.

6 Conclusion

In this paper, a novel compressive sensing-base method for face recognition has been proposed. We used the CS-2FFT feature extraction and SRC classifiers to make our proposed face recognition scheme become less computationally complex and more robust than some existing face recognition algorithms. To this end, first we extracted CS-2FFT feature vectors from images and then we used these feature vectors on the SRC classifier to classify face images. Experiment results showed that the CS-2FFT feature extraction and the SRC classification are effective and robust in the face recognition task. In particular, our proposed face recognition scheme achieves 96% accuracy on the ORL dataset, 95.9% accuracy on the Extended Yale B face database dataset, 95.6% accuracy on the AR database dataset and 42% accuracy on the LFW dataset with a less computational complexity when compared to other state-of-the-art face recognition approaches.

References

1. R. Henson and R. Cohen Kadosh and M. Johnson and F. Dick, *Task-dependent Activation of Face-sensitive Cortex: An fMRI Adaptation Study*, IEEE Journal of Cognitive Neuroscience, 2010, 22, 5, 903 - 917, May.
2. A. Martinez and R. Benavente, *The AR face database*, CVC Tech. Report., 1998, 1, 24, April.
3. K. Lee, and J. Ho, and D. Kriegman, *Acquiring linear subspaces for face recognition under variable lighting*, IEEE Trans. on Pattern Analysis and Machine Intelligence., 2003, 27, 5, 684-698, April.
4. J. Ju, and K. N. Plataniotis, and A. N. Venetsanopoulos, *Regularized discriminant analysis for the small sample size problem in face recognition*, Pattern Recognition Letters., 2003, 24, 16, 3079-3087, April.
5. P. J. Grother, and R. J. Micheals, and P. J. Phillips, Face Recognition Vendor Test 2002 Performance Metrices, *Proc. of Int. Conference Audio- and Video-Based Biometric Person Authentication*, 937-945, 2003, June.
6. I. Loris, *On the performance of algorithms for the minimization of l_1 -penalized functionals*, Inverse Problems, 2009, 25, 3, 1-16, April.
7. P. N. Belhumeur, and J. P. Hephner, and D. J. Kriegman, *Eigenfaces vs. Fisherfaces: recognition using class specific linear projection*, IEEE Trans. on Pattern Analysis and Machine Intelligence, 1997, 19, 7, 711-720, April.
8. M. Turk, and A. Pentland, *Eigenfaces for recognition*, Journal of Cognitive Neuroscience, 1991, 3, 1, 71-86, May.
9. E. Learned-Miller, A. Ferencz and F. Jurie, *Faces in Real-Life Images*, Proc. Faces in Real-Life Images workshop at the European Conference on Computer Vision, 5-10, 2008.

10. P. Viola and M. Jones, *Robust Real-time Object Detection*, International Journal of Computer Vision, 2001, 4, 3, 1-5, May.
11. B. S. Manjunath and R. Chellappa and C. von der Malsburg, *A feature based approach to face recognition*, Proc. IEEE Conference on Computer Vision and Pattern Recognition, pages = 373-378, 1992.
12. M. Lades and J. C. Vorbruggen and J. Buhmann and J. Lange and C. von der Malsburg and R. P. Wurtz and W. Konen, *Distortion invariant object recognition in the dynamic link architecture*, IEEE Trans. on Computers, 1993, 42, 3, 300-311, March.
13. R. Chellappa and C. L. Wilson and S. Sirohey, *Human and machine recognition of faces: A survey*, Proceedings of The IEEE, 1995, 83, 5,705-740, May.
14. S. Lawrence and C. L. Giles and A. C. Tsoi and A. D. Back, *Face recognition: a convolutional neural-network approach*, IEEE Trans. on Neural Networks, 1997, 8, 1, 98-113, Jan..
15. J. Daugman, *Face and gesture recognition: Overview*, IEEE Trans. on Pattern Analysis and Machine Intelligence, 1997, 19, 7,675-676, July.
16. L. Wiskott and J.-M. Fellous and N. Kuiger and C. von der Malsburg, *Face recognition by elastic bunch graph matching*, IEEE Trans. on Pattern Analysis and Machine Intelligence, 1997, 19, 7,775-779, July.
17. K.-M. Lam and H. Yan, *An analytic-to-holistic approach for face recognition based on a single frontal view*, IEEE Trans. on Pattern Analysis and Machine Intelligence, 1998, 20, 7, 673-686, July.
18. H. E. Komleh and V. Chandran and S. Sridharan, *Robustness to expression variations in fractal-based face recognition*, Proc. IEEE International Symposium on Signal Processing and its Applications, 359-362, 2001.
19. A. M. Martinez, *Recognizing imprecisely localized, partially occluded, and expression variant faces from a single sample per class*, IEEE Trans. on Pattern Analysis and Machine Intelligence, 2002, 24, 6, 748-763, June.
20. T. Ojala and M. Pietikainen and T. Maenpaa, *Multiresolution gray-scale and rotation invariant texture classification with local binary patterns*, IEEE Trans. on Pattern Analysis and Machine Intelligence, 2002, 24, 7, 971-987, July.
21. J. Wu and Z.-H. Zhou, Face recognition with one training image per person, Pattern Recognition Letters, 2002, 23, 14, 1711-1719, July.
22. B. Kepenekci and F. B. Tek and G. B. Akar, *Occluded face recognition based on Gabor wavelets*, Proc. IEEE International Conference on Image Processing, 373-378, 2002..
23. W. Zhao and R. Chellappa and P. J. Phillips and A. Rosenfeld, *Face recognition: A literature survey*, ACM Computing Surveys, 2003, 35, 4, 399-458, Dec..
24. J. Huang and P. C. Yuen and W.-S. Chen and J. H. Lai, *Component-based LDA method for face recognition with one training sample*, Proc. IEEE International Workshop on Analysis and Modeling of Faces and Gestures, 2003, 35, 4, 120-126, Oct., Nice, France.
25. H.-C. Jung and B.-W. Hwang and S.-W. Lee, *Authenticating corrupted face image based on noise model*, Proc. IEEE International Conference on Automatic Face and Gesture Recognition, 272-277, 2004.
26. J. Yang and D. Zhang and A. F. Frangi and J.-U. Yang, *Two-dimensional PCA: a new approach to appearance-based face representation and recognition*, IEEE Trans. on Pattern Analysis and Machine Intelligence, 2004, 26, 1, 131-137, Jan.
27. S. Chen and J. Liu and Z.-H. Zhou, *Making FLDA applicable to face recognition with one sample per person*, Pattern Recognition, 2004, 37, 7, 1553-1555, July.
28. H.-S. Le and H. Li, *Recognizing frontal face images using hidden Markov models with one training image per person*, Proc. IEEE International Conference on Pattern Recognition (ICPR'04), 318-321,2004.
29. M. H. Shaheed, *Performance analysis of 4 types of conjugate gradient algorithms in the nonlinear dynamic modelling of a TRMS using feedforward neural networks*, Proc. IEEE International Conference on Systems, Man and Cybernetics, 5985-5990, 2004, The Hague, Netherlands, Oct.
30. X. Tan and S. Chen and Z.-H. Zhou and F. Zhang, *Recognizing partially occluded, expression variant faces from single training image per person with SOM and soft KNN ensemble*, IEEE Trans. on Neural Networks, 2005, 16, 4, 875-886, July.
31. Y. Gao and Y. Qi, *Robust visual similarity retrieval in single model face databases*, Pattern Recognition, 2005, 38, 7, 1009-1020, July.

-
- 32. X. Tan and S. Chen and Z.-H. Zhou and F. Zhang, *Face recognition from a single image per person: A survey*, Pattern Recognition, 2006, 39, 9, 1725-1745, Sept.
 - 33. Y. Zhang and A. M. Martinez, *A weighted probabilistic approach to face recognition from multiple images and video sequences*, ELSEVIER Image and Vision Computing, 2006, 24, 6, 626-638, June.
 - 34. G. Huang and M. Ramesh and T. Berg and E. Learned-Miller, *Labeled faces in the wild*, 1-17, 2013, 2, May.
 - 35. S. Sheel and T. Varshney and R. Varshney, *Accelerated learning in MLP using adaptive learning rate with momentum coefficient*, Proc. IEEE International Conference on Industrial and Information Systems, 307-310, 2007.
 - 36. M. R. Every, *Discriminating Between Pitched Sources in Music Audio*, IEEE Trans. on Audio, Speech, and Language Processing, 2008, 16, 2, 267-277, Feb.
 - 37. P. Hennings-Yeomans and S. Baker and B. V. K. V. Kumar, *Recognition of Low-Resolution Faces Using Multiple Still Images and Multiple Cameras*, Proc. IEEE International Conference on Biometrics: Theory, Systems, and Applications, 56-61, 2008, Sept.
 - 38. T. V. Sreenivas and W. B. Kleijn, *Compressive sensing for sparsely excited speech signals*, Proc. IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP), 4125-4128, 2009.
 - 39. M. Wakin, *Compressive sensing*, 4125-4128, 2009, Sept.
 - 40. P. Nagesh and B. Li, *A Compressive Sensing Approach for Expression-Invariant Face Recognition*, Proc. IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 1518-1525, 2009.
 - 41. X. Yap and A. W. H. Khong and W.-S. Gan, *Localization of acoustic source on solids: A linear predictive coding based algorithm for location template matching*, Proc. IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP), 2490-2493, 2010, March, Dallas, USA.
 - 42. F. Paulin and A. Santhakumaran, *Classification of breast cancer by comparing back propagation training algorithms*, International Journal on Computer Science and Engineering, 2011, 3, 1, 327-332, Jan.
 - 43. S. Liang and Y. Wang and Y. Liu, *Face Recognition Algorithm Based on Compressive Sensing and SRC*, Proc. IEEE International Conference on Instrumentation, Measurement, Computer, Communication and Control (IMCCC), 1460-1463, 2012, Dec., Harbin City, China.
 - 44. D.-Y. Chen and P.-C. Hsieh, *Face-based Gender Recognition Using compressive sampling*, Proc. IEEE International Symposium on Intelligent Signal Processing and Communications Systems (ISPACS), 157-161, 2012, Nov., New Taipei.
 - 45. M. Banitalebi-Dehkordi and H. R. Abutalebi and M. R. Taban, *Sound source localization using compressive sensing-based feature extraction and spatial sparsity*, Digital Signal Processing Elsevier, 2013, 23, 4, 1239-1246, July.
 - 46. M. Banitalebi-Dehkordi and A. Banitalebi-Dehkordi, *Music Genre Classification Using Spectral Analysis and Sparse Representation of the Signals*, Journal of Signal Processing Systems, 74, 2, 273-280, 2014.
 - 47. Hind Chebbo et al., *The ORL Database of Faces*, 1-17, 2013, 2, May, Myconos, Greece, Invited Paper), available online in <http://www.cl.cam.ac.uk/research/dtg/attarchive/facedatabase.html>, May 2013.
 - 48. A Banitalebi, M Moosaei, and GA Hossein-Zadeh, An investigation on the usage of image quality assessment in visual speech recognition, 3rd International Congress on Image and Signal Processing (CISP), 2010 China.
 - 49. D. Tao, J. Cheng, X. Gao, X. Li, and Ch. Deng, Robust Sparse Coding for Mobile Image Labeling on the Cloud, IEEE Transactions on Circuits and Systems for Video Technology, Volume: 27, Issue: 1, Jan. 2017.
 - 50. D. Tao, Y. Guo, M. Song, Y. Li, Zh. Yu, and Y. Y. Tang, Person Re-Identification by Dual-Regularized KISS Metric Learning, IEEE Transactions on Image Processing, Volume: 25, Issue: 6, June 2016.
 - 51. Edited by: M. Kawulok, M. Emre Celebi, and B. Smolka, Labeled Faces in the Wild: A Survey. In Advances in Face Detection and Facial Image Analysis, Springer, pages 189-248, 2016.

-
- 52. Y. Sun, X. Wang, and X. Tang, Deep convolutional network cascade for facial point detection, In Computer Vision and Pattern Recognition (CVPR), 2013 IEEE Conference on, pages 3476-3483. IEEE, 2013.
 - 53. Y. Sun, X. Wang, and X. Tang, Hybrid deep learning for face verification, In Computer Vision (ICCV), 2013 IEEE International Conference on, pages 1489-1496. IEEE, 2013.
 - 54. Y. Sun, X. Wang, and X. Tang, Deep learning face representation from predicting 10,000 classes, In Computer Vision and Pattern Recognition (CVPR), 2014 IEEE Conference on, pages 1891-1898. IEEE, 2014.
 - 55. Y. Taigman, M. Yang, M. Ranzato, and L. Wolf, Deepface: Closing the gap to human-level performance in face verification, In Computer Vision and Pattern Recognition (CVPR), 2014 IEEE Conference on, pages 1701-1708. IEEE, 2014.
 - 56. E. Zhou, Z. Cao, and Q. Yin, Naive-deep face recognition: Touching the limit of LFW benchmark or not?, arXiv preprint arXiv:1501.04690, 2015.
 - 57. Vitomir Struc, PhD (Pretty helpful Development) functions for face recognition toolbox, available at: <http://www.face-rec.org/source-codes/>, 2012.