Sparse representation for face recognition: A review paper

Jitendra Madarkar | Poonam Sharma | Rimjhim Padam Singh

Computer Science and Engineering, VNIT, Nagpur, India

Correspondence
Poonam Sharma, Computer Science and Engineering, VNIT, Nagpur, Maharashtra, India. Email: jitendramadarkar475@gmail.com

Funding information
Ministry of Electronics and Information technology, Grant/Award Number: MLA/MUM/GA/10(37)B

Abstract
With the increasing use of surveillance cameras, face recognition is being studied by many researchers for security purposes. Although high accuracy has been achieved for frontal faces, the existing methods have shown poor performance for occluded and corrupt images. Recently, sparse representation based classification (SRC) has shown the state-of-the-art result in face recognition on corrupt and occluded face images. Several researchers have developed extended SRC methods in the last decade. This paper mainly focuses on SRC and its extended methods of face recognition. SRC methods have been compared on the basis of five issues of face recognition such as linear variation, non-linear variation, undersampled, pose variation, and low resolution. Detailed analysis of SRC methods for issues of face recognition have been discussed based on experimental results and execution time. Finally, the limitation of SRC methods have been listed to help the researchers to extend the work of existing methods to resolve the unsolved issues.

1 | INTRODUCTION
Since the last two decades, face recognition (FR) is playing a very significant role in authentication and security purpose. Recently, several homes and companies have face recognition based biometric system for security purpose. The world lurks under the threat of terrorism and the need of the hour is a good system of human face recognition, so that it recognizes the human face from the images and video. Since, the researchers have been working in the area of face recognition and they have achieved state-of-the-art results for frontal face images.

Face recognition is mainly a two-step process: first step is feature extraction and the second is classification. Discriminative feature extraction from the face image is a difficult task due to occlusion, noise, disguise, variation, and similar structure of the faces. Several methods have been proposed to extract discriminative features from the face image like principal component analysis (PCA) [1], local binary pattern LBP [2], Fisher discriminant analysis (FDA) [3], independent component analysis (ICA) [4], and locality preserving projections (LPP) [5], curvelet [6]. The aforementioned feature extraction methods are known to support the linear data and not non-linear data. The non-linear data can be converted into a linear data by converting original space into a high-dimensional space with the help of kernel trick method or kernel mapping function like kernel PCA (KPCA) [7], kernel FDA (KFDA) [7], and support vector machine (SVM) [8]. After the feature extraction process, these extracted features are used for classification to recognize a label of probe sample from the training dataset. Classification plays an important role to find out a correct class for the probe sample from the training dataset. There are several existing classifiers which are used for face recognition such as nearest neighbour (NN) [9], nearest feature line (NFL) [10], hidden Markov model (HMM) [11], nearest feature plane (NFP) [12], nearest feature subspace (NFS) [12], SVM [8], linear regression based classification (LRC) [13], and sparse representation based classification (SRC) [14]. There exists numerous challenges in face recognition such as linear variation, pose variation, rotation, non-linear variation, undersampled, scaling, and low resolution [15]. These issues degrade the performance of face recognition methods and they failed to solve all the issues of face recognition. Since a decade, SRC has shown good performance over all the existing methods on an occluded face image [14, 16, 17]. Initially, sparse representation was used in signal processing for compression [18] and representation rather than inference and classification. The main purpose of sparse representation was to produce the most compressed signal with the help of a linear combination of atoms in an overcomplete dictionary [19]. Many researchers have proved that the sparse representation is an extension of traditional methods like Fourier, wavelet and...
is being used widely in the field of reconstruction of signal.SRC method has been used in many practical applications of the computer vision such as face recognition [14], image super-resolution [20], denoising [21], clustering [22], image segmentation [23], image deblurring [24], image classification [25]. For the very first time Wright et al. [14] introduced the SRC method on face recognition in 2009 and since then SRC has been used in face recognition application with some modifications. SRC has shown better performance and more discriminating power than existing classifier. NN classifies a probe sample into the given training class that contains the closest neighbour sample. NN uses a single sample from the class for classification. NN classifier degrades the performance of face recognition due to outlier or noise. NFL uses two samples to express the probe sample from the same class. NFL enhances the express capacity of the classifier by passing the line through each pair of the samples belonging to the same class. NFL enhances the representation capacity of the probe sample by passing the plane through three samples from the same class. NN, NFL, and NFL are not efficient when the classes have a number of samples. In NFS, the probe sample approximates with all samples from the same class and predicts a class label which produces minimum reconstruction error. SVM is mostly used for linear classification but for non-linear classification kernel method is needed. It is difficult to choose SVM parameters and appropriate kernel method. SVM also needs a feature extraction method to extract features from the image. SRC represents a probe sample by a linear combination of dictionary atoms, which means the samples from all classes and a probe sample belonging to the minimum residual producing class. SRC does not care about the feature of input data and it needs a large number of samples for each class. It has achieved state-of-the-art results in face recognition application. Wright et al. applied SRC on noisy, occlusion, and disguise images. In [14], authors have achieved better performance for face recognition and it has been concluded that SRC tolerates small pose variation. SRC is a global representation that never considers locality of the feature during classification. Its performance depends on a sparse coefficient and dictionary. Researchers have developed several extended methods of SRC by modifying dictionary and sparse coefficient vector. Initially, a dictionary was constructed by using the training samples, but due to variation in probe sample, it cannot be well represented by dictionary atoms. There are many popular dictionary learning algorithms [26] which update the dictionary atoms with proper prototypes such as $k$ singular value decomposition (KSVD) [27], label consistent KSVD (LC-KSVD) [28], Fisher discrimination dictionary learning (FDDL) [29], D-KSVD, support vector guided dictionary learning (SVGDL) [30], and DLSPC-G. The linear equation used in sparse representation can be solved by convex relaxation method [31] and greedy method [32].

The main aim of the survey paper is to find the strength and limitation of SRC methods, which helps the researchers to extend the work of existing methods to resolve the unsolved issues.

The paper is organized as follows. In Section 2, the SRC framework has been introduced. Section 3 discusses five different categories of SRC and its extended methods for face recognition. The different extended SRC (ESRC) methods have been compared based on their performance of face recognition and this has been provided in Section 4. Section 5 presents experimental result. Section 6 discusses conclusion and future aspects.

## 2 Sparse Representation Based Classification

SRC finds the sparsest representation of a probe sample from the dictionary which is constructed by all the training samples and classifies a probe sample to the class which produces less residual error.

Let $A = [A_1, A_2, ..., A_N] \in \mathbb{R}^{\text{w} \times \text{d}}$ be the set of training samples of all classes are called as dictionary. Each dictionary column is referred to as an atom or a basis. There are total $n$ number of atoms in the dictionary. $k$ represents number of classes and let $A_k = [a_{k1}, a_{k2}, ..., a_{kn}]$ represent the set of samples that belongs to the $i$th class. $N$ is number of samples in a class.

The linear Equation (1) has multiple solutions, $\ell_0$-minimization, or a greedy method such as matching pursuit (MP) [33], orthogonal matching pursuit (OMP) [34] algorithms are used to find optimized solution as shown in Equation (2).

$$y = Ax \in \mathbb{R}^n,$$

where $x = [0, 0, 0, \alpha_{i1}, \alpha_{i2}, ..., \alpha_{in}, 0, 0, 0]^T$.

The linear Equation (1) finds the sparsest representation of probe sample $y$, but in case of undetermined dictionary, Equation (1) has multiple solutions, $\ell_0$-minimization, or a greedy method such as matching pursuit (MP) [33], orthogonal matching pursuit (OMP) [34] algorithms are used to find optimized solution as shown in Equation (2).

$$\hat{x} = \arg \min_{\alpha} \|x\|_0 \; \text{s.t.} \; y = Ax,$$

where $\| \cdot \|_0$ refer to as $\ell_0$-norm which counts non-zero entries in a vector. To solve a linear Equation (2) by $\ell_0$-minimization is an NP-hard problem where it is difficult to find an approximation. If a solution of $x_0$ is sparse enough, then it can be solved by using $\ell_1$-minimization in polynomial time [35]. Then Equation (2) can be modified with the help of $\ell_1$-minimization or convex relaxation methods [31] as shown in Equation (3).

$$\hat{x} = \arg \min_{\alpha} \|x\|_1 \; \text{s.t.} \; y = Ax,$$

$$\hat{x} = \arg \min_{\alpha} \|x\|_1 \; \text{s.t.} \; y - Ax \leq \varepsilon,$$

where $\| \cdot \|_1$ refers to $\ell_1$-norm which performs summation of all in $X$ vector. When the images are captured in an unconstrained environment it may be affected by noise and hence, Equation (3) is not enough to find the proper approximation for noisy
images. Then the linear Equation (3) can be extended for noisy or corrupted images with the help of error or noise vector as shown in Equation (4). $z \in \mathbb{R}^m$

$$y = Ax + z.$$  

(4)

When the probe sample is partially occluded or corrupted then the linear model that is defined in Equation (2) can be extended for noisy or corrupted images with the help of error or noise vector as shown in Equation (4). $z \in \mathbb{R}^m$

$$y = Ax + z.$$  

(4)

The basic idea of SRC is to search representative atoms from the dictionary which is constructed by all the training samples that can sparsely represent a probe sample $y$ and its sparse coefficient is obtained by solving the convex problem of the linear system using Equations (3) and (5). A probe sample is assigned to the class that minimizes the residual error. SRC assumes that the samples from the same class lie on same subspace. Hence, probe sample $y$ is classified based on minimal residual or the class which has a more number of non-zero entries in sparse coefficient vector $x \in \mathbb{R}^s$, the residual is computed by Equation (6).

$$\min_{r_i(y)} = \|y - Ax_i\|^2,$$  

(6)

where $r_i(y)$ is a residual of probe sample $y$ with respect to the $i$th class.

Non-face images can be validated or classified separately using SRC. If non-zero coefficients are spread across all over the classes of sparse vector $x_i$ then that probe sample is neither from gallery set nor a face image at all. Thus it can be classified separately as no face image. The SRC has been performing better under specific condition only. These conditions are:

- SRC needs the assumption that all samples from single class lie on same subspace.
- Subspace from one class should not intersect with other classes.
- Each class has enough number of samples.
- Training samples should not be corrupted.
- Training samples contain more variation and pose variation.

Several ESRC methods have been proposed to overcome the above-mentioned SRC issues.

### 3 | BROAD CATEGORIES OF SRC BASED FACE RECOGNITION ALGORITHMS

When the image is captured in an unconstrained environment, it may be possible to have linear variation, non-linear variation, pose variation, and low-resolution issues. There are other issues of face recognition due to less number of samples, lack of variation in training samples, and corrupted/noisy training samples. In this survey, SRC and its extended methods have been broadly classified for different problems of face recognition as follows:

#### 3.1 | Linear variation for face images

Illumination, occlusion, and corruption degrade the performance of face recognition. SRC [14] has overcome the issues of linear variation and performed better than existing holistic methods like PCA, LDA, and ICA. As SRC is robust to occlusion [14], it uses occlusion dictionary to solve the problem of occlusion in face recognition but occlusion matrix size is very big and it is observed that big dictionary increases the time complexity. Yang [36] proposed Gabor feature based SRC (GSRC) method where the authors used Gabor feature based occlusion dictionary instead of occlusion dictionary to reduce the size of occlusion dictionary. Gabor features are invariant to constrained illumination, rotation, expression, and scaling. These features help GSRC to improve FR performance. The GSRC represents test sample by modifying (5) in Equation (7).

$$G(y) = \left(G(A) \ G(I)\right) \begin{bmatrix} x \\ \epsilon \end{bmatrix},$$  

(7)

where $G(A)$ is a dictionary which contain Gabor feature and $G(I)$ is a Gabor feature based occlusion dictionary.

The GSRC method used Gabor features of training samples as a dictionary atoms whereas SRC method used training samples as a dictionary atoms. SRC requires the assumption that the training samples of each class follows linearity (globally linear) or lies on the same subspace. Sometimes subspace of samples of any class can intersect the subspace of other classes. In such situations, the probe sample can be represented by dictionary atoms of other classes, though the training sample of correct class lies on the same subspace. SRC still requires another assumption that different classes subspace should not intersect with each other, that is SRC does not support when the data is globally non-linear. It is noted that if the data is globally non-linear then SRC can misclassify the probe sample. Hui et al. [35] proposed sparse representation based classification SNRC method, that finds first $k$ basis from each class which are the nearest neighbours to probe sample $y$, whereas SRC finds a basis from all the classes. First, $k$-nearest neighbour samples of each class form the local
dictionary and SNRC classifies the class which can be best approximate with the probe sample $y$.

$$x_i = \arg \min_{x} \| \psi x - y \|_2 - \lambda \cdot y^T : x + y : \| x \|_1,$$  \hspace{1cm} (8)

where, $\psi$ is a dictionary that contains first $k$ nearest samples of $y$ in class $i$. In SNRC, the authors first used KNN method to build local dictionary but that is not appropriate on complex image data or corrupted face images. It is hence said that it cannot select the correct nearest neighbour samples for the probe sample $y$ from each class of training dataset. After they applied SRC on non-appropriate selected samples or atoms to approximate sparse coefficient of probe sample $y$, it is thus observed that some correct samples of training data could not be part of the local dictionary. The inappropriate local dictionary will not harness sparsity, hence it could misclassify probe sample. The modular approach in SRC [14] divides the face image into different parts or modules and uses majority voting for classification. In this approach, occluded and clean model takes part equally in final classification decision. But in case of equally weighted, clean module loses correlation between other modules which is most important for final classification. To overcome these issues, Lai et al. [38] proposed weighted global sparse representation (WGSR) method that considers only clean models for classification and utilizes correlation information of clean modules in the recognition process. The reliability of module is obtained by sparsity and the weight of the module is determined by the residual of the module. The weight of the module reconstructs global feature vector that helps the final classification. This method only works for frontal images and affects the performance if images have pose and non-linear variation. SRC fails to provide good performance, when the face image data distribution is non-linear. Non-separable linear data can be converted into linear separable data by using kernel trick or non-linear mapping function. Non-linear mapping function transfers data from original feature (original dimension) space $\mathbb{R}^m$ into the high-dimensional space $H_m \rightarrow \mathbb{R}^m$. Yin et al. [39] proposed kernel based SRC called KSRC, in this method a mapping function was proposed to convert original feature (original dimension) space into high-dimensional feature space and then apply SRC on the high-dimensional feature space. If input images are noiseless then the computational cost of SRC and KSRC are same. If the dimension of the sample is more than a number of samples in each class then KSRC works better than SRC, contrary to this KSRC is less efficient than SRC. Let $A = [\phi(A_1), \phi(A_2), ..., \phi(A_k)]$ represent dictionary of high-dimensional training samples. To approximate sparse coefficient, KSRC used $\ell^1$-minimization to solve the convex problem by Equation (9).

$$x_i = \arg \min_{x} \| x \|_1 \ s.t. \ \| \phi (y) - Ax \|_2 \leq \epsilon.$$ \hspace{1cm} (9)

It is an improvement of SRC but the selection of the kernel function for different images is a difficult task and still poses as an open challenge for the researchers. KSRC method increases the dimension of the data, hence the time complexity increases and fails to resolve the issues of SRC. SRC decides the class label of probe sample $y$ based on the minimum class residual. Li et al. [16] suggested that the residual based decision rule is not efficient for sparsity. Jiang proposed new decision rule for SRC based on the sum of the coefficient. Sparse coefficient measures the similarity between a probe sample and the training samples. Maximum sum of coefficient (SoC) easily decides the class label of probe sample by Equation (10).

$$\text{identi}f(y) = \arg \min_{i} \ (1')^T x', \ i = 1, 2, ..., k,$$ \hspace{1cm} (10)

where, $1'$ is a vector or can we say identity vector, $i$ indicates class label. Here sparse coefficient value is directly proportional to the similarity. SoC based SRC performs better than residual based SRC on ideal images. If the probe sample is partially corrupted, occluded, and affected by illumination then it is observed that SoC based SRC method produces less coefficient with correct class or produces more coefficient with incorrect class. Hence, it may generate less similarity with samples of correct class. In low-dimensional or downsampled image, subspaces have very less discriminative information and hence linearity structure is not enough for the SRC. When linearity does not support for classification that time locality can do a better job in classification. Lu et al. [40] proposed a weighted sparse representation based classification (WSRC) method that enforces the locality restraint on the sparsity regularized reconstruction problem. In this scenario, SRC generates different sparse coefficient vector even for similar probe samples using $\ell^1$-minimization method. It is hence noted that, the sparse coefficient classifies the probe sample by training samples which may belong to the incorrect class, that leads to the unreliable classification result. WSRC method unites both data locality and sparsity structure into expression (11) to overcome the downsampled problem. While sparse linear approximation WSRC conserves the similarity between probe sample and neighbouring training samples. The WSRC approximate sparse vector by Equation (11) using $\ell^1$-minimization method.

$$\tilde{x}_i = \arg \min_{x} \| Wx \|_1 \ s.t. \ Ax = y,$$ \hspace{1cm} (11)

where, $W$ is the locality adaptor that calculates the distance between probe sample $y$ and training samples by Equation (12).

$$\text{diag}(W') = [\text{dist} (y, x_1'), ..., \text{dist} (y, x_m')]^T$$

$$\text{dist} (y, x') = \| y - x' \|_1,$$ \hspace{1cm} (12)

where, $i$ is a locality adaptor parameter if $i=0$, then WSRC works similar to SRC, $c$ represents number of class, and $m$ is a total sample. The same person having different variation samples produces different distances using Equation (12). Distance measure cannot work when the samples have a different variation in it. WKSRC [41] is a combination of KSRC [39] and WSRC [40], by combining these methods the authors have tried to overcome the issues of locality and non-linearity. KSRC solved the linearity but does not carry the locality in feature data. Similarly,
WSRC supports locality but does not support non-linearity. Liu et al. [41] proposed WKSRC, this method first transfers the data from low-dimensional space to high-dimensional space using non-linear mapping function or kernel trick and then applies dimensionality reduction method such as KPCA and KFDA on high-dimensional feature space to reduce the cost of computation and storage space, after this it applies weighted matrix to sustain locality in data. Illumination is a big challenge in face recognition application, but illumination variation is overcome by applying the MSR algorithm. This method obtains efficient coefficient vector by combining kernel feature space and locality information but did not solve the issues of SRC. According to Hu et al. [42], dictionary has a significant role in SRC, but has a challenge in finding relationship between class label and dictionary atoms due to the variation in training and probe samples. In simple SRC, the dictionary model exploits the classification on the basis of the residual and sparse coefficient that restricts the performance of face recognition. To overcome this issue, Hu et al. [42] introduced a new novel dictionary building method, where dictionary was constructed by a common basis and discriminative basis. Apply PCA on a matrix of training data $A$ and select $p$ number of eigenvector, which is shared by all individuals of training sample called molecular dictionary $G$ and training data itself act as an atom dictionary $A$ which is discriminative in nature. Then, the extended dictionary is incurred by the merging of the atom dictionary and molecular dictionary. Extended dictionary $D=[A,G] \in \mathbb{R}^{n \times (s+p)}$ is replaced by dictionary in Equation (13).

$$\hat{x} = \underset{\hat{\beta}}{\text{arg min}} \frac{1}{2} \| A \hat{x} - y \|_2^2 + \lambda \| \hat{x} \|_1, \quad \text{subject to} \quad \| A \hat{x} - y \|_2^2 \leq \varepsilon. \tag{13}$$

Non-zero elements of the sparse vector tell the similarity between probe sample and dictionary atom. Any image finds more similarity with the image itself but produces less similarity with Eigenvector of all training samples. Hence there is no valid reason for merging the molecular dictionary with the atom dictionary.

Lu et al. [43] proposed the FDLSR method which replace $\ell^1$ by $\ell^{2-1}$-norm, which takes advantage of both the norms. $\ell^1$-norm is discriminative in nature and $\ell^2$-norm uses for systematic representation. The $\ell^{2-1}$-minimization is used to decrease the negative impact of an outlier and it is performed to select class-specific samples across all the data points with some sparsity. This method shows stronger discriminant power than the other representation models because dictionary is learned by Fisher discrimination criteria. The $\ell^{2-1}$-norm is robust to the noise but this method does not tackle the issue of unconstrained condition. The $\ell^{2-1}$-norm can be generalized as shown in Equation (14).

$$\| A \|^{2-1}_{2-1} = \sum_{i=1}^{n} \| \varphi_i \|_2^2, \tag{14}$$

where $A$ is a matrix and $a$ is a vector. The WSCR [38] method considers the same weights for all the patches and it ignores the different variations in patches. The different variations can change the weight of the patches. Ma et al. [44] introduced a new method robust SRC in an adaptive weighted spatial pyramid structure (R SRC-ASP) which makes the use of sparsity evaluation based self-adaptive weighting strategy for residual aggregation, which provide different weights with respect to the variation of the patches. This algorithm needs proper alignment images and it is not possible practically. In R SRC-ASP, the final aggregated residual estimated by Equation (15).

$$r_{ij}(y) = \sum_{k} \frac{\| y_k - A^k \alpha_k^i \|_2^2}{\| \alpha_k^i \|_1^2}, \tag{15}$$

where $\frac{1}{\| \alpha_k^i \|_1}$ is the self-adaptive weight corresponding to sub-patch $k$, $\alpha_k^i$ is the sub-sparse coefficient vector over the $i$th class dictionary atom of patch $k$.

Nowadays deep learning has shown tremendous performance in computer vision to extract the automatic features from the image. The authors Chang et al. [63] have used stacked convolutional autoencoder (S C AE) to extract the features to construct a dictionary and then applied SRC for classification. The deep learning method extracts complex feature whereas other feature extraction methods do not extract the complex features. Thus this method is able to solve linear variation and illumination but does not solve the pose and the undersampled issues of FR.

The resolution of an image plays a significant role in the performance of FR and resolution need to change for betterment of algorithm. Author Jitendra proposed a new method multi-resolution sparse representation based classification (MRSRC) [67], this method used compress wavelet features as a dictionary atom but this method is not effective for non-linear variation. SRC, SNRC, KSRC, WSR, SoC based SRC, MRSRC, and SRC-S C AE degrade the performance of face recognition on non-linear variation, pose variation, and undersampled issues.

### 3.2 Non-linear variation for face images

Expression, smile, and local deformation comes under non-linear variation. The SRC uses the holistic feature in the dictionary that cannot handle expression and local deformation effectively. If the probe sample contains non-linear variation, then a single dictionary is not enough for the classification. An extra dictionary needs to be added that contains non-linear variation or any mechanism that can find a proper prototype for each class so that the dictionary can represent non-linear variation also. In the GSR C [36], the Gabor feature is less sensitive to non-linear variation, hence the Gabor feature of training samples is used as a dictionary atom. The GSR C has shown an improvement over SRC. Nevertheless, SRC needs uncorrupted and more variation training samples to resolve non-linear variation. The ESRC [17] has overcome non-linear variation issues.
Authors use two different extended $\pi$ type membership functions for the generation of extended $\pi$ interval type-II membership function, as shown in Equation (17).

$$\tilde{\pi} = \left[ \pi^L; \pi^U \right],$$  \hspace{1cm} (17)

where subscripts $L$ and $U$ represent lower and upper membership functions, respectively.

The quality of any algorithm depends on the accuracy and time complexity of that algorithm. The accuracy of FR depends on features of the image and time complexity is based on the time required to execute the algorithm. The authors have considered the aforementioned issues and solved these issues by developing a new method which used Gabor features that are invariant to illumination, expression, and rotation. The least number of features can give high results and faster execution hence authors have used supervised locality preserving projections (SLPP) to reduce the dimension of features and this method used fast matching pursuit (FMP) method to represent a probe sample which is faster than homotopy algorithm which is used in other SRC based methods. The authors have combined all the advantages and proposed Gabor-SLPP-FMP (GSF) framework [61] which is efficient than conventional SRC methods. This method does not solve undersampled, pose issues of FR.

Existing methods had used global features and local features to resolve FR issues. But it is difficult to solve FR issues by single global features or local features. The authors have used both the features with addition to statistic property and logarithmic operation. The authors have proposed LWS for feature descriptor and fast extended sparse-weighted representation classifier (FESWRC) for the representation of a probe sample. The logarithmic-weighted sum (LWS) feature descriptor combines the significance of local features (DOST), global features (Gabor), a statistical property of image (covariance), and logarithmic operation. The FESWRC used the primal augmented Lagrangian method (PALM) for fast execution. Finally, FDWLSR [65] is obtained by combining LWS and FESWRC. This method can solve linear and non-linear variation with the help of the aforementioned preprocessing method but is unable to solve undersampled, pose low-resolution issues. The FESWRC can be explained by Equation (18).

$$y = W_{A^4} \otimes A x_1 + W_{D_f^4} \otimes D_f \tilde{\beta} + \tilde{\gamma},$$  \hspace{1cm} (18)

where $W_{A^4}$ and $W_{D_f^4}$ are weight coefficient vector of corresponding dictionaries $A$ and $D_f$, respectively. The symbol $\otimes$ represents the Hadamard product operator.

The sparsity constrained $\|x\|_1$ and fidelity $\|y - Ax\|_2^2$ are two important terms in Equation (3). The $\ell^2$-norm is used as fidelity term and this term is impacted by a large error caused due to more variation in samples. When the corrupted pixels is more then the linear Equation (3) may fail. The Mi [68] have proposed a new method, robust supervised sparse representation (RSSR), by improving fidelity term and collaborative...
representation strategy is helpful to improve recognition accuracy. This method uses Huber loss as fidelity term in linear equation.

The RSSR uses two-phase representation scheme to implement supervised sparse representation. The first phase, referred to as the coarse representation, uses all training samples to represent a probe; while in the second phase, referred to as a fine supervised sparse representation, only training samples which have high contribution in the coarse representation are used and the coefficients for the rest of the training samples are set to zero.

3.3 Undersampled for face images

SRC has shown prominent results when a large number of samples are available in each class, but in the case of undersampled, SRC has performed worst. If the training dataset has less number of samples in each class and samples are noisy, then the individual class does not have an appropriate prototype for the classification. Hence, the probe sample fails to sparsely represent a correct class due to noise in training samples. To overcome this issue of undersampled, Deng [17] proposed a new ESRC method. In the ESRC method, authors have introduced an extra intraclass dictionary that is created by subtracting the prototype or natural image of a class from the samples of the same class. The linear representation of Equation (4) is modified by introducing intraclass dictionary \( D \) and is shown in Equation (19).

\[
y = Ax_1 + D_f \beta + \varepsilon,
\]

where, \( D_f \) is the intraclass variant matrix that represents the local expression, environmental illumination, and occlusion of the face image, where these changes cannot be modelled by a small dense noise matrix \( \varepsilon \) in Equation (4). The ESRC method represents a probe sample using two dictionaries \( A \) and \( D_f \) by Equation (20).

\[
\begin{align*}
\left( \begin{array}{c}
x_1 \\
\beta_f
\end{array} \right) &= \arg \min_{\beta} \| \left( \begin{array}{c}
x \\
\beta
\end{array} \right) \|_1 \\
\text{s.t.} \quad & \| (A \ D_f) \left( \begin{array}{c}
x \\
\beta
\end{array} \right) - y \|_2 \leq \varepsilon.
\end{align*}
\]

In ESRC, the first dictionary takes part in the classification process of probe sample and the second dictionary is used to generate a more stable sparse vector. The extra variant dictionary is constructed from the variation of the samples of the same class but when within the class, variation samples are found to be less, then probe sample classifies the other class that has more variation samples. The second dictionary of the ESRC method does not take part in the classification. Each class has a different variation which means some classes have less variation in samples and the first dictionary also contains the same variation hence ESRC reduces efficiency. The above-mentioned problem having less variation is solved by the same author with the help of a new method called a superposed SRC method [48]. In this method, the first part of the dictionary atoms of each class is replaced by a single prototype or centroid of a class. However, this method does not provide a reliable classification due to the single prototype or centroid of each class. The centroid of each class is used to represent and determine the label of probe sample. This single prototype or centroid can lead to misclassification in case of variation. Zhang et al. [49] proposed the OSPP-SR method, which is an enhancement to the ESRC method by improving the intravariant dictionary. The OSPP-SR method required a lesser number of atoms in the intravariant dictionary, resulting in reduced computational cost. Iliadis et al. [50] combine sparsity approach of ESRC with least square and proposed a new method, sparse representation, and regularized least squares (SR + RLS) that first finds sparse coefficient \( x \) using ESRC Equation (20) and then constructs a small dictionary with non-zero dictionary atom and finally, approximate \( f \) using regularized least square (21).

\[
\hat{f} = \arg \min_f \| y - \hat{T} f \|_2^2 + \lambda \| f \|_2.
\]

In this method, the authors use two classification mechanisms to find a correct class instead of using any learning mechanism for a proper prototype of the respective classes. The SR_RLS method has improved performance over ESRC but fails to overcome the issues of ESRC. The ESRC cannot solve the problem of corrupted training samples and a few variations containing samples of the individual class. These issues are resolved by Jiang who proposed the SDR-SLR [45] method that represents probe sample by a collaboration of class-specific dictionary and the dense combination of the non-class-specific dictionary (intraclass variation dictionary). The ESRC uses the variation to represent a probe sample from the same class, whereas SDR takes the collaborative variation of different classes. The ESRC represents the probe sample through training samples using Equation (20). In SDR-SLR, first-class-specific components are extracted from natural images and then it is by SRC as shown in Equation (22).

\[
y = a + b + s.
\]

Where \( a \) is a class-specific component and represented by dictionary \( A \) that contains class-specific component using linear model \( a = A\alpha + e_a \), \( b \) is a non-class-specific component and it is represented by dictionary \( B \) that uses linear model \( b = B\beta + e_b \), \( s \) indicate random sparse noise subspace or noise matrix. The SDR represents the probe sample that is shown in Equation (23). It uses two dictionaries and SDR-SLR approximates their sparse coefficient using Equation (24).

\[
y = A\alpha + B\beta + e \quad e = e_a + e_b + s,
\]

\[
\min_{\alpha, \beta, e} \|\alpha\|_1 + \|\beta\|_2^2 + \lambda \| e \|_1 \quad \text{s.t.} \quad y = A\alpha + B\beta + e.
\]

The SSRC requires more number of samples to get a single centroid or prototype of the class. The centroid or prototype of a class is estimated by performing a linear operation like
average. That does not give discriminative information for classification. To overcome the issue of non-linear variation in single labelled sample per person (SLSPP), Gao et al. [46] proposed an $S^3RC$ method which used the ESRC framework with modification of the gallery dictionary. $S^3RC$ used two dictionaries: gallery dictionary is estimated by Equation (26) and variation dictionary is estimated by Equation (26). The intravariate dictionary helps to rectify linear variation like illumination, occlusion, and the gallery dictionary is used to reduce non-linear variation like expression, smile. ESRC and SSRC methods did not learn a gallery dictionary to resolve the non-linear issue. $S^3RC$ used GMM to find a single prototype for each class using labelled and unlabelled samples. The GMM learns prototype or centroid using a semi-supervised EM algorithm. The performance of FR is depends on the result of GMM-EM but data needed to follow normal distribution. The $S^3RC$ represent a probe sample by Equation (27).

$$V = [A_t - e_1T, ..., A_k - e_kT]$$ or (25)

$$V = [G_t - e_1T, ..., G_k - e_kT] \quad \text{(SLSPP problem)}$$

$$P = [c_1, ..., c_k]$$ or (26)

$$P = [T] \quad \text{(SLSPP problem)}$$

$$\left(\hat{\alpha}, \hat{\beta}\right) = \arg \min_{\alpha, \beta} \|P \ V \ \left(\hat{\alpha}, \hat{\beta}\right) - y\|^2_2$$

$$+ \lambda \|\hat{\alpha}, \hat{\beta}\|_1, \quad \text{(27)}$$

where $P$ is a gallery dictionary and $V$ is variation dictionary.

The image features are complex structures (non-linear), but simple SRC methods do not solve the non-linear structure. But kernel methods can convert the non-linear structure into a linear structure. This problem can be solved by the KSRC method in sparse representation. Fan et al. have taken advantage of this method and proposed a method called kernel coordinate descent based on virtual dictionary (KCDVD) [51], which constructs the virtual dictionary by approximately symmetrical face images (ASFI) by using Equations (28) and (29). Due to the virtual dictionary, it alleviates the problem of undersampled but the KCDVD method does not solve the other issue of KSRC except undersampled.

$$l_1^{t+1} = l_1^t - \xi (l_1^t - l_1), \quad \text{(28)}$$

$$l_2^{t+1} = l_2^t - \xi (l_2^t - l_2), \quad \text{(29)}$$

where $\xi$ is the learning rate and $t$ represents the iteration or time, $l_1$ is a left part of the face image (column vector) similarly $l_2$ is a right part of the face image (column vector). The main motive of ASFI method is to reduce the error of $\|l_1 - l_2\|^2$, which is done by gradient descent scheme. The KCDVD method generates a virtual sample from the same image but is not able to add different variations in the newly created images. Thus this method can only solve linear variation problem but is not able to solve non-linear variation, and pose variation issues of FR.

The undersampled issue was solved by authors [62] by developing a random filter sparse representation (RFSR) method. This method follows the two-step scheme. In the first step create new virtual random filtering images from original training samples using Equation (30) and create a new training database by combining newly created samples with original samples. In the second step perform CRC on the newly created training dataset. Authors also add noise in newly create virtual images. But realistic variations are totally different than creating random filtering images, hence it could not solve the realistic variation problem.

$$A^c_{i-j}(p, q) = \frac{1}{3} \sum_{\substack{p \leq c \leq p+3 \pm 4 \pm w}} A^c_{i-j}(p, q), \quad \text{(30)}$$

where $p$ and $q$ are the $X - Y$ coordinates of an image $A$, $s$ is a scale of filter $L = t * t$, $t$ is always an odd number and decides the size of the filter.

When the training sample is insufficient then many existing methods have developed invariant dictionary to improve the performance of FR. But authors proposed new a method transfer learning based sparse representation and weighted fusion (TLSRWF) to tackle undersampled issue where it use transfer learning method and weighted fusion scheme for sparse representation. The residual is estimated by Equations (31) and (35).

$$d_1^t = \|y - A(\hat{x})\|_2, \quad \text{(31)}$$

$$Z = \hat{x} = [\hat{x}_1, \hat{x}_2, ..., \hat{x}_n], \quad \text{(32)}$$

$$t = Z\beta = [Z_1, Z_2, ..., Z_n] \beta = \hat{\beta}_1, \hat{\beta}_2, ..., \hat{\beta}_n, \quad \text{(33)}$$

$$\hat{\beta} = \arg \min_{\beta} \|t - Z\beta\|^2_2 + \|\beta\|^2_2, \quad \text{(34)}$$

$$d_2^t = \|t - Z(\hat{\beta})\|_2, \quad \text{(35)}$$

We combined the weights from two different reconstruction errors and finally classified test samples by evaluating which class leads to the minimum weighted fusion

$$f_\xi = w_1d_1^t + w_2d_2^t. \quad \text{(36)}$$

If $kk = \text{argmin} f_\xi$, then the test sample is finally assigned to the $k$th class.
The SDR-SLR extract class-specific information using single value decomposition (SVD) but when the training images are corrupted by non-Gaussian noise then it is hard for SDR-SLR to obtain optimal low-rank matrix. The low-rank component contains important facial features and are beneficial for better representation. A low-rank recovery is a key technique that can separate corrupted information from the training face images. Yang et al. [52] proposed a new sparse low-rank component based representation (SLCR) method based on low-rank recovery. It constructs the dictionary by low-rank component and non-low-rank component which are approximate by using low-rank matrix recovery as shown in Equation (37), hence this method is better than the other SRC methods for the undersampled and low-quality images. The linear equation of SLCR is represented by Equation (38) and a sparse coefficient is estimated by Equation (39).

\[ D = L + N, \]  
\[ y = L\alpha + N\beta + \varepsilon, \]  
\[ \min_{\alpha,\beta,\varepsilon} \|\alpha\|_1 + \gamma\|\beta\|_2^2 + \lambda\|\varepsilon\|_1 \quad s.t. \quad y = L\alpha + N\beta + \varepsilon, \]  
where \( y \) and \( \lambda \) are constants for a compromise.

The residual is estimated by Equation (40)

\[ r(y) = \|\varepsilon - \zeta\|_2 = \|L(I - C_i)\alpha\|_2, \]  
where \( I \) is an identity matrix, \( L \) is the low-rank dictionary of SLCR, \( C_i \) is a class label matrix of the training dataset \( D \) for class \( i \), its element \( C_i(k, k) = 1 \) if the \( k \)th training image originates from class \( i \) and all other elements of \( C_i \) are zero.

The ESRC [17] method extract variation from the samples of the same class and KSRC [39] method transfer low dimensional space data into high-dimensional space. The Fan et al. combine the concept of KSRC with ESRC and proposed fast kernel ESRC method which takes advantage of both the methods but does not alleviate the issue of non-linear variation when the samples of a class do not have different variation. The FKESRC achieved high computational efficiency by using the coordinate descent method in the feature space.

### 3.4 Pose variation for face images

Pose variation is a big issue in face recognition, when the probe sample has pose variation with training samples then face recognition performance degrades rapidly. Pose variation changes the features of the face image with respect to their position, so these features can make a strong correlation with samples of other classes having the same pose rather than a sample of the correct class. To overcome this issue, Feilong et al. [53] proposed the multi-weighted sparse representation (MW-SR) method. In case of pose variation, a similar pose shares more contribution and if the sample is from the same person then it contributes the most. In this method, first all training samples are separated into \( p \) group on the basis of poses, \( A_j \in \mathbb{R}^{m \times n}, i = 1, 2, \ldots, \mathcal{K} \) and order the direction of the face from left pose to right pose. The dictionary \( A \) and a probe sample \( y \) are given, the sparse vector \( x \) is estimate by solving the linear Equation (3) then compute residual by using Equation (6). The minimum residual decides the closest pose group of a probe sample \( y \). Once the pose of the probe sample is obtained then all weights for all pose with respect to probe sample are calculated. Weights are calculated by Equation (41) and the authors assume that weights followed a Gaussian distribution.

\[ w_i = e^{-\Pi_{i(i-1)}}, \quad i = 1, 2, \ldots, \mathcal{K} \]  
where, \( u \) is the sub-index of the determined closest pose by SR method and \( s \) is the sampling step. There may be a chance of selecting a different pose sample, thus finding appropriate similar pose weights has been introduced into SRC. The MW-SR model can extend by \( \ell^1 \)-minimization as shown in Equation (42).

\[ \tilde{x}_1 = \arg\min_{x} \sum_{i=1}^{p} \|x\|_1 \quad s.t. \quad \sum_{i=1}^{p} w_i A_i x_i = y. \]  
In the MW-SR method, weights build the relation between different poses of the face. Even if the training dataset does not have a similar pose like a probe sample, still this method is able to classify or identify a probe sample. The MW-SR method needs more training samples with different poses, but in real-world scenarios, it may not be possible to obtain different pose images for all training classes. In the case of undersampled and non-linear variation, the MW-SR method degrades the performance drastically.

When the training and testing samples are captured in a different environment then the distribution of the images might be different due to unconstrained variations such as pose, illumination, and expression. For better result distribution mismatch needs to be reduced. The authors [64] proposed a new method subspace extended sparse representation classifier and learning discriminative feature SESRC-LDF to tackle the above face recognition problem. This method has used two classification strategies on the basis of probe-sample symmetry. The symmetry of a sample has small variation then it will be recognized by SESRC, otherwise it will be recognized by LDF. Before solving sparse linear objective function, the goal of subspace learning is to find an optimal transformation matrix that transforms all testing and training samples into a common subspace as shown in Equation (43). The transform matrix is used to reduce the intra-class differences and enlarge the inter-class differences.

\[ \hat{y}^+ = \pi^T Y^+ = [\hat{y}_1, \hat{y}_2, \ldots, \hat{y}_3], \]  
\[ \tilde{y}^+ = \pi^T \tilde{y}^+ = [\tilde{\pi}B], \]  
where \( A \) represents a texture structure gallery dictionary and \( B \) is a texture structure invariant dictionary. A \( \pi \) is a transform matrix that is used to learn discriminative subspace.
3.5 Low resolution for face images

Low-resolution (LR) images have less discriminative features for image classification, hence LR images degrade the performance of face recognition. When the training dataset has high-resolution images and a probe sample is a low-resolution image then there is a mismatch of resolution and discriminative feature. The major challenge of low-resolution face images to be resolved are: first, convert LR images into high-resolution (HR) images, second HR training images convert into LR images, and third both LR and HR images are embedded into common subspace. There are several ways to tackle the low-resolution problem of face recognition like super-resolution (SR), interpolation. Bilgazyev et al. [54] have shown improvement using the super-resolution technique in face recognition. Initially, Yang et al. [20] proposed super-resolution (SR) using sparse representation, for the construction of HR images from LR images. In [55, 56], authors have used SR using sparse representation in face recognition and shown improved performance.

The SLCR construct the dictionary by low-rank component and non-low-rank component which are approximate by using low-rank matrix recovery, hence this method is better for low-quality images.

4 COMPARISON OF DIFFERENT SRC METHODS

In this section, a comparison of SRC and all its extended methods have been done on five problems of face recognition.

- Linear variation: linear variation contains illumination effect, noise, and occlusion.
- Non-linear variation: local expression or deformation.
- Undersampled: less number of samples in each class of training dataset.
- Pose variation: should not be more than 20 degrees.
- Low resolution: very few pixels contain the whole image information.

All SRC methods have been compared on the basis of their significant advantages, disadvantages, and learning mechanism as used in SRC. All ESRC methods are developed by modification of sparse coefficients, dictionary, and fusing new methods into SRC. The SoC SRC, WSRC, MW-SR, WKSRC are developed by modifying the sparse coefficient. The GSRC, ESRC, SSRC, SR-RLS, SDR-SLR, RFSR, molecular, KCDVD, CSR, $S^3$RC, TLSRWF, FDWLSR, ExIntTy2KBSRM, SESRC-LDF, RSRC-ASP, SLCR, and MRSRC are developed by modifying dictionary and SNRC, KSRC, WKSRC, Gabor-SLPP-FMP (GSF), SRC-SCAE, FKESRC methods are developed by fusing existing methods into SRC.

From Table 1 it is clear that all SRC methods are able to solve linear variations, but to solve non-linear variation requires one more intraclass variation dictionary or needs a learning mechanism like GMM, arithmetic operation to find proper centroid or prototype of the class.

The methods which do not used any extra dictionary to estimate sparse coefficient, these methods are able to solve the only linear issue such as SRC, WGSR, KSRC, SNRC, SoC SRC, WSRC, molecular, SRC-SCAE, WKSRC, MRSRC. The methods that are able to solve linear as well as a non-linear issues such as GSRC, ESRC, SSRC, SR-RLS, SDR-SLR, $S^3$RC, SLCR, CSR, TLSRWF, Gabor-SLPP-FMP (GSF), RFSR, FDWLSR, RSRC-ASP, ExIntTy2KBSRM, FKESRC failed to overcome the issues of low resolution and pose variation ($>20^\circ$). The methods which are used to create new dictionary with the help of virtual samples or probe samples are able to solve undersampled issue such as ESRC, SSRC, SR-RLS, SDR-SLR, $S^3$RC, KCDVD, RSRC-ASP, TLSRWF, RFSR, SLCR, FKESRC. A pose variation issue is solved by the MW-SR and SESRC-LDF methods by keeping all pose variation images in training samples but in real life scenario to gather all types of pose images is not possible. Till date, no such SRC method is developed to overcome the issue of low resolution. The SDR, SLCR, and $S^3$RC are the best-performing methods among all SRC methods and are used to solve the issue of linear variation, non-linear variation, and undersampled.

The ESRC, SSRC used one more intravariant dictionary and extract variation from other samples using arithmetic operation hence these methods are able to solve non-linear variation and undersampled.

The $S^3$RC methods find a single prototype for each class using labelled and unlabelled samples. This method uses GMM and the parameter learns by EM method which is able to solve linear variation, non-linear variation, and undersampled issues of FR. The SRC-SCAE method is able to solve linear variation and illumination due to deep learning features. The deep learning technique required a large number of samples with different variations but is not possible for all classes, hence this method solves only linear variation. The ExIntTy2KBSRM method combined fuzzy logic with KSRC. The membership function fuzzify the face database and it helps to measures the unseen information present in the pixels. This method solves linear and non-linear variations. The Gabor-SLPP-FMP (GSF) framework uses the Gabor feature which is invariant to non-linear variation, the same as GSRC but it is faster than GSRC methods. This method solves linear and non-linear variations. The FDWLSR has used global features (Gabor), local features, statistical property, and logarithmic operation, these feature are invariant to linear and non-linear issues of FR. The KCDVD method generates a virtual sample from the same image and is not able to add different variations in the newly created images, hence it solves the linear and undersampled issue of FR. The RFSR method creates virtual random filtering images and adds noise in the newly created image but realistic variations are totally different than creating random filtering images. This method also solves the linear and undersampled issue of FR. The SLCR method is based on low-rank recovery, the low-rank component contains important facial features and is beneficial for better representation. The FKESRC method solved undersampled and linear variation issues but ESRC uses arithmetic operation to extract variation from other samples of the same class which is not effective for non-linear variation. The RSRC method used Huber loss.
as a fidelity term and two-phase representation scheme which alleviates the non-linear issue of FR. The class-specific component features are better for all variations except pose variation. The MRSRC method uses compressed wavelet features as a dictionary, thus it only solves linear variation. The SESRC-LDF method has uses two classification strategies on the basis of probe-sample symmetry. If the symmetry of a sample has small variation then it will be recognized by SESRC, otherwise it will be recognized by LDF. Before solving sparse linear objective function, the goal of subspace learning is to find an optimal transformation matrix that transforms all testing and training samples into a common subspace. The transform matrix is used to reduce the intraclass differences and enlarge the interclass differences. Till date, no such SRC method has been developed to overcome the issue of low resolution. The SDR-SLR, SLRC, and $S^3$RC are the best-performing methods among all SRC methods and are used to solve the issue of linear variation, non-linear variation, and undersampled.

## EXPERIMENTAL RESULTS

In this section, different state-of-the-art SRCs and its extended methods of face recognition have been evaluated. All SRC methods have done experimentation on face databases given in Table 2. AR and Extended Yale B (EYB) are the most used databases in the experimentation of SRC and its extended methods. Here, we have performed experimentation on different methods using six databases such as AR [50], EYB, labeled faces in the wild (LFW) [51], CMU PIE [59] georgia tech (GT), and olivetti research laboratory (ORL).

### 5.1 Experiment setup

In this subsection, the experimental details of different SRC methods is provided. The regularized parameter was set 0.001 for all methods and homotopy algorithm was used to represent

| Sr. No. | Methods          | Linear variation | Non-linear variation | Undersampled | Pose variation | Low resolution |
|---------|------------------|------------------|----------------------|--------------|----------------|----------------|
| 1       | SRC (2009)       | Y                | N                    | N            | N              | N              |
| 2       | GSRC (2010)      | Y                | Y                    | N            | N              | N              |
| 3       | WGSRC (2012)     | Y                | N                    | N            | N              | N              |
| 4       | KSRC (2012)      | Y                | N                    | N            | N              | N              |
| 5       | SNRC (2012)      | Y                | N                    | N            | N              | N              |
| 6       | ESRC (2012)      | Y                | N                    | N            | N              | N              |
| 7       | SoC SRC (2013)   | Y                | N                    | N            | N              | N              |
| 8       | WSRRC (2013)     | Y                | N                    | N            | N              | N              |
| 9       | SSRC (2013)      | Y                | N                    | N            | N              | N              |
| 10      | SR-RLS (2014)    | Y                | N                    | Y            | N              | N              |
| 11      | SDR-SLR (2015)   | Y                | Y                    | Y            | N              | N              |
| 12      | MW-SR (2015)     | Y                | N                    | N            | Y              | Y              |
| 13      | WKSRC (2016)     | Y                | N                    | N            | N              | N              |
| 14      | Molecular (2016) | Y                | N                    | N            | N              | N              |
| 15      | FDLSR (2016)     | Y                | N                    | N            | N              | N              |
| 16      | CSR (2016)       | Y                | Y                    | N            | N              | N              |
| 17      | S3RC (2017)      | Y                | Y                    | Y            | N              | N              |
| 18      | KCDVD (2017)     | Y                | N                    | Y            | Y              | N              |
| 19      | ExIncTy2KBSRM (2018) | Y                | Y                    | N            | N              | N              |
| 20      | TLSRWF (2018)    | Y                | N                    | Y            | N              | N              |
| 21      | RSR-ASP (2018)   | Y                | N                    | Y            | N              | N              |
| 22      | SRC-SCAE (2018)  | Y                | N                    | N            | N              | N              |
| 23      | RFSSR (2018)     | Y                | N                    | Y            | N              | N              |
| 24      | GSF (2018)       | Y                | Y                    | N            | N              | N              |
| 25      | FDWLSR (2019)    | Y                | Y                    | N            | Y              | N              |
| 26      | SLCR (2019)      | Y                | Y                    | Y            | N              | N              |
| 27      | SESRC-LDF (2019) | Y                | N                    | N            | N              | Y              |
| 28      | FKESRC (2020)    | Y                | Y                    | Y            | N              | N              |
| 29      | MRSRC (2020)     | Y                | N                    | N            | N              | N              |
probe sample in sparse vector. MATLAB 2016 was used for the experimentation having the configuration of 8 GB RAM, 3.6 GH i-7 processor.

5.2 Face databases

5.2.1 AR database

AR [50] face database contains 126 individuals having 26 color samples of each individual. Images are frontal view faces with different illumination condition, occlusion, and facial expression. Images were captured in the constraint condition. Samples of an individual are shown in Figure 1.

5.2.2 Extended Yale B

EYB face database contains 38 individuals and each individual having 9 different poses. Each pose has 64 different illumination samples. Samples of an individual are shown in Figure 2.

5.2.3 LFW (Labeled faces in the Wild) face database

The LFW [51] face database images that are taken from the web of 1680 individuals. These images are varying in illumination, pose, expression, and different background. LFW database is a benchmark for face recognition application. Samples of an individual are shown in Figure 3.

5.2.4 CMU PIE database

CMU PIE face database contains samples with different variation such as illumination, expression, and pose. The database samples were collected in Carnegie Mellon University in 2000. Samples of an individual are shown in Figure 4.
5.2.5 GT (Georgia Tech) face database

In GT face database contains images of 50 people and each individual has 15 samples. The images were taken in Georgia Institute of technology for image processing. The images have frontal, tilted view with different facial expression and illumination effect. Samples of an individual has been shown in Figure 5.

5.2.6 ORL (Olivetti Research Laboratory) face database

ORL face database contains total 400 face images of 40 individuals. Each individual sample was captured at the different time, expression, and lighting condition. All images have the same background. Samples of an individual are shown in Figure 6.
5.3 Experimental results on different issue

Here, we have setup experimentation on different issues of face recognition such as illumination, occlusion, expression, undersampled, and low resolution.

5.3.1 Illumination

AR database
We have taken 100 individuals and each individual has 12 samples. We choose six neutral samples per individual as training dataset and six illuminated samples per individual as testing dataset. The experimental result is shown in Figure 7.

CMU PIE database
We have taken 30 individuals and each individual has 21 samples varying with different illumination effect. We randomly selects 2, 3, 4, 5, 7, 9, and 11 number of samples for training and rest of the samples for testing. The experimental results are shown in Figure 8.

EYB database
We have taken 30 individuals and each individual has 20 samples varying with different illumination effects. We randomly selects 2, 5, 10, 15 samples for training and rest of the samples for testing. Both the sets have illuminated variation samples. The experimental result is shown in Figure 9.

5.3.2 Occlusion

EYB database
We have taken 100 individuals and each individual has 12 samples. We choose six neutral samples per individual as training dataset and six occluded samples per individual as a testing dataset. The experimental result is shown in Figure 10.

The low-rank approximation is performing better for occlusion. Thus SDR performs well in case of the occlusion.
5.3.3 | Expression

**AR database**

We have taken 100 individuals and each individual has 14 samples. We choose (1) 4 neutral samples and (2) 10 neutral with illumination samples per individual as training dataset. Choose four samples with expression per individual as testing dataset. The experimental result is shown in Figure 11.

**Undersampled AR database**

We have taken 100 individuals and each individual has 14 samples with the size of $44 \times 44$. We choose randomly 2, 3, 4, 5, 6, 7, 8, and 9 samples per individual as training dataset and rest of the samples for testing dataset. The experimental result varies based on number of training samples and is shown in Figure 12.

**EYB database**

We have taken 38 individuals and each individual has 16 samples with the size of $32 \times 32$. We choose randomly 2, 3, 4, 5, 6, 7, 8, and 9 samples per individual as training dataset and rest of the samples for testing dataset. The experimental result varies based on number of training samples and is shown in Figure 13.

**LFW database**

We have taken 34 individuals and each individual has 9 samples with the size of $50 \times 50$. We choose randomly 2, 3, 4, 5, 6, 7, and 8 samples per individual as training dataset and rest of the samples for testing dataset. The experimental result varies based on number of training samples and is shown in Table 3.

In case of undersampled issue, low-rank component and intravariant dictionary are performed better hence, SDR ESRC are performed better.

**GT database**
We have taken 50 individuals and each having 15 samples. We randomly choose 2, 3, 4, 5, 6, 7, 8, 9 samples per individual as a training dataset and use all the others for testing with 63 × 42 size. The experimental result varies based on number of training samples and it is shown in Figure 14.

**ORL database**
We have taken 40 individuals and each individual has 10 samples with the size of 44 × 36. We choose randomly 2, 3, 4, 5, 6, 7, and 8 samples per individual as training dataset and rest of the samples for testing dataset. The experimental result varies based on number of training samples and is shown in Figure 15.

### 5.3.5 Low resolution

**LFW database**
We have taken 35 individuals and each individual has 9 samples. We choose randomly five samples per individual as a training and rest of the samples for testing. Dimension of the images were resized from 10 × 11 to 90 × 90. The experimental result is shown in Figure 16.

**ORL database**
We have taken 40 individuals and each having 10 samples. We randomly choose five samples per individual as a training dataset and used all the others for testing. Dimensions of the images are resized from 9 × 11 to 92 × 112. The experimental result varies based on resolution of samples and it is shown in Figure 17.

The result of different methods SRC, GSRC, RRC, ESRC, WSRC, SSRC, SR-RLS, SDR-SLR, CSR, and S^3RC on different databases AR [50], EYB, LFW [51], CMU PIE, GT, and ORL. As labelled samples decrease the performance of all SRC methods also decreases.
5.4 Analysis of execution time

The SRC and its extended methods are expensive with respect to the execution time. In this paper, the comparison of different methods with respect to execution time such as SRC, GSRC, ESRC, SSRC, SR-RLS, SDR-SLR, and CSR have been provided. In SRC based methods, the computational cost is directly proportional to the size of the dictionary. The execution time of SRC and its extended methods are given in Table 4. The SR-RLS method consumes less execution time and the order of execution time is followed by SRC, SSRC, ESRC, GSRC, SDR-SLR, and CSR. The SR-RLS method, that constructs a new dictionary on the basis of sparse coefficient estimated by ESRC which suppresses many training samples to zero, hence SR-RLS reduces the execution time. The ESRC constructs an extra variant dictionary from the samples of the same class and uses two dictionaries to calculate sparse vector, hence the execution time increases for ESRC. Whereas in case of SSRC, the first part of dictionary atoms of each class are replaced by single prototype, which reduces the size of the dictionary, and improves the time efficiency of SSRC above ESRC. The GSRC method has used occlusion dictionary which increases time complexity. The SDR-SLR is the slowest, it consumes time to decompose images into the class-specific component, nonclass-specific component, and needs a lot of optimization process to extract proper component. The CSR method requires more time to execute. Table 4 has shown the time consumption of SRC and its extended methods on five different databases such as AR, EYB, LFW, ORL, and GT. For experimentation, the dataset is divided into training images set and testing images set randomly. All experimental setup and image sizes are the same as mentioned in the undersampled Section 5.3.4

5.5 Analysis of experimental result

The experimental setup has been categorized on the basis of five issues such as illumination, occlusion, expression, low resolution, and undersampled. The experimental design helps to analyze the results better. The following observations have been observed:

1. When the training dataset contains neutral samples and test dataset contain same pose with illumination and expression variation sample then ESRC and SSRC performed better.
2. If the training dataset contains neutral samples and test dataset contain same pose with occlusion variation then SDR performed better because low-rank approximation is efficient for occlusion.
3. If both training and testing have similar variation then SDR-SLR performed better.
4. Low-resolution images can reduced the performance for all SRC methods.
5. In undersampled case, if the extra variance dictionary is not used then the performance of SRC methods decrease rapidly.
6. From the Figures 14–18 it can deducted that when the training samples for each individual increase then the performance also increases.
7. In case of one sample per person (OSPP), SDR-SLR performs better in all the experimental databases but in LFW it has been shown to be giving less accuracy.
8. SR-RLS is faster among all the methods and SDR-SLR method is slowest among all the methods.
9. Sparse coefficient value indicates the similarity between train and test sample. The similar images have high sparse coefficient value as shown in Figures 18 and 19. In SRC, the

| Methods | AR Accuracy | AR Time | EYB Accuracy | EYB Time | ORL Accuracy | ORL Time | LFW Accuracy | LFW Time | GT Accuracy | GT Time |
|---------|-------------|---------|--------------|---------|--------------|---------|--------------|---------|-------------|--------|
| SRC     | 85.0        | 20.5    | 94.6         | 5.6     | 94.4         | 1.5     | 45.5         | 2.8     | 70.8        | 13.3   |
| GSRC    | 78.64       | 664     | 94.3         | 88      | 98.5         | 66      | 42.4         | 32      | 53.07       | 210    |
| ESRC    | 92.8        | 96.2    | 94.8         | 16.7    | 95.9         | 4.3     | 47.7         | 4.2     | 74.3        | 29.1   |
| SSRC    | 92.4        | 38.0    | 95.5         | 9.8     | 96.6         | 3.2     | 46.7         | 3.4     | 74.2        | 20.4   |
| SR-RLS  | 92.3        | 27.2    | 95.5         | 5.0     | 95.4         | 1.0     | 39.1         | 0.5     | 69.2        | 7.1    |
| SDR-SLR | 93.4        | 870.8   | 95.3         | 143.5   | 94.8         | 58.5    | 56.8         | 48.1    | 73.8        | 289.3  |
| CSR     | 92.7        | 4326    | 94.3         | 897     | 94.1         | 260     | 58.0         | 214     | 68.7        | 1775   |
test sample predicts the class which gives minimum residual. By observing the bar chart as shown in Figures 18 and 19, we can conclude that the class which has the highest coefficient value is not enough to classify the correct class. SRC needs more number of samples which can give high sparse coefficient and produce minimum residual for a test sample. In some scenario, SRC may misclassify when with the correct class have more different variation and less similar samples. Figure 18 shows sparse coefficient value of LFW dataset where samples have more variation and Figure 19 shows sparse coefficient value of AR dataset, where samples have less variation. By observing both the figures, we conclude that the less variation training sample have better representation than more variation samples.

6 | CONCLUSION

In this paper, we have provided a review of SRC and its extended methods for face recognition, which are developed by modifying sparse coefficient and dictionary. The different variants of SRC have been classified on the basis of their capability of solving the issues of face recognition such as linear variation, non-linear variation, undersampled, pose variation, and low resolution. The comparison of all variants of SRC methods and have seen their advantages and disadvantages with respect to issues of face recognition, have been provided. It is observed that some methods are able to solve linear variation, non-linear variation, and undersampled problems. It is also observed that SRC had not fully explored the issue of low resolution and pose variation. We have also provided the experimental results of various SRC and its extended methods for different issues of face recognition with different numbers of training and testing samples of the face databases such as AR, EYB, LFW, CMU PIE, GT, and ORL.

The result analysis of SRC along with its extended methods have been discussed on the basis of their experimental result. It is observed that SDR-SLR, CSR, ESRC, and SSRC perform better in non-linear variation and undersampled case. It is concluded that SR + RLS performs better when samples contain less variation. As we decrease the number of labelled samples performance also decreases, but in case of OSPP $S^3RC$ method performs better as compared to all other undersampled methods.

In the future, SRC may be developed in all its respect to explore the issues of low resolution and pose variation in face recognition.

ACKNOWLEDGEMENTS

The work was supported by Visvesvaraya PhD scheme, Government of India.
FUNDING
This study was funded by the Ministry of Electronics and Information Technology (India) (Grant No.: MLA/MUM/GA/10(37)B).

CONFLICT OF INTEREST
The authors declare that there are no conflicts of interest regarding the publication of this paper.

REFERENCES
1. Matthew, T., Alex, P.: Eigenfaces for recognition. J. Cogn. Neurosci. 3, 71–86 (1991)
2. Ahonen, T., Hadid, A., Pietikäinen, M.: Face description with local binary patterns: Application to face recognition. IEEE Trans. Pattern Anal. Mach. Intell. 28, 2037–2041 (2006)
3. Belhumeur, P.N., Hespanha, J.P., Kriegman, D.J.: Eigenfaces vs. fisherfaces: Recognition using class specific linear projection. IEEE Trans. Pattern Anal. Mach. Intell. 19, 711–720 (1997)
4. Bartlett, M.S., Movellan, J.R., Sejnowski, T.J.: Face recognition by independent component analysis. IEEE Trans. Neural Netw. 13, 1450–1464 (2002)
5. Xiaoifei, H., et al.: Face recognition using Laplacian faces. IEEE Trans. Pattern Anal. Mach. Intell. 27, 328–340 (2005)
6. Poonam, S., Yadav, R., Arya, K.: Pose-invariant face recognition using curvelet neural network. IET Biometrics 3, 128–138 (2014)
7. Yang, M.: Kernel Eigenfaces vs. kernel fisherfaces: Face recognition using kernel methods. In: Proceedings of Fifth IEEE International Conference on Automatic Face Gesture Recognition, pp. 215–220 (2002)
8. Phillips, P.J.: Support vector machines applied to face recognition. In: Proceedings of the 1998 Conference on Advances in Neural Information Processing Systems II, pp. 803–809 (1999)
9. Cover, T., Hart, P.: Nearest neighbor pattern classification. IEEE Trans. Inf. Theory 13, 21–27 (1967)
10. Li, S.Z., Lu, J.: Face recognition using the nearest feature line method. IEEE Trans. Neural Netw. 10, 439–443 (1999)
11. Nefian, A.V., Hayes, M.H.: Face detection and recognition using hidden Markov models. In: Proceedings of 1998 International Conference on Image Processing. ICIP’98 (Cat. No.98CH36269) 1, pp. 141–145 (1998)
12. Chien, J.-T., Wu, C.-C.: Discriminant waveletfaces and nearest feature classifiers for face recognition. IEEE Trans. Pattern Anal. Mach. Intell. 24, 1644–1649 (2002)
13. Naseem, I., Togneri, R., Bennamoun, M.: Linear regression for face recognition. IEEE Trans. Pattern Anal. Mach. Intell. 32, 2106–2112 (2010)
14. Wright, J., et al.: Robust face recognition via sparse representation. IEEE Trans. Pattern Anal. Mach. Intell. 31, 210–227 (2009)
15. Zhifei, W., et al.: Low-resolution face recognition: A review. Vis. Comput. 30, 359–386 (2014)
16. Jiang, L., Can-Yi, L.: A new decision rule for sparse representation based classification for face recognition. Neurocomputing 116, 265–271 (2013)
17. Deng, W., Hu, J., Guo, J.: Extended SRC: Undersampled face recognition via intraclass variant dictionary. IEEE Trans. Pattern Anal. Mach. Intell. 34, 1864–1870 (2012)
18. Donoho, D.L.: Compressed sensing. IEEE Trans. Inf. Theory 52, 1289–1306 (2006)
19. Huang, K., Aiyi, S.: Sparse representation for signal classification. Adv. Neural Inform. Process. Syst. 19, 609–616 (2007)
20. Yang, J., et al.: Image super-resolution via sparse representation. IEEE Trans. Image Process. 19, 2861–2873 (2010)
21. Giryes, R., Elad, M.: Sparsity-based Poisson denoising with dictionary learning. IEEE Trans. Image Process. 23, 5057–5069 (2014)
22. Chen, Y., et al.: In-plane rotation and scale invariant clustering using dictionaries. IEEE Trans. Image Process. 22, 2166–2180 (2013)
23. Cao, K., Liu, E., Jain, A.K.: Segmentation and enhancement of latent fingerprints: A coarse to fine ridge structure dictionary. IEEE Trans. Pattern Anal. Mach. Intell. 36, 1847–1859 (2014)
24. Xiang, S., et al.: Image deblurring with coupled dictionary learning: A coarse to fine ridge structure dictionary. Int. J. Comput. Vis. 114, 248–271 (2015)
25. Zhang, Z., et al.: Improved sparse representation method for image classification. IET Computer Vision 11, 319–330 (2017)
26. Zhang, Z., et al.: A survey of sparse representation: Algorithms and applications. IEEE Access 3, 490–530 (2015)
27. Aharon, M., Elad, M., Bruckstein, A.: K-SVD: An algorithm for designing overcomplete dictionaries for sparse representation. IEEE Trans. Signal Process. 54, 4311–4322 (2006)
28. Jiang, Z., Lin, Z., Davis, L.S.: Label consistent K-SVD: Learning a discriminative dictionary for recognition. IEEE Trans. Pattern Anal. Mach. Intell. 35, 2651–2664 (2013)
29. Yang, M., et al.: Fisher discrimination dictionary learning for sparse representation. In: 2011 International Conference on Computer Vision, pp. 543–550 (2011)
30. Cai, S., et al.: Support vector guided dictionary learning. Comput. Vis.-ECCV 8692, 624–639 (2014)
31. Tropp, J.A.: Algorithms for simultaneous sparse approximation. Part II: Convex relaxation. Signal Process. 86, 589–602 (2006)
32. Joel, T., Anna, G., Martin, S.: Algorithms for simultaneous sparse approximation. Part I: Greedily pursuit. Signal Process. 86, 572–588 (2006)
33. Mallat, S.G., Zhang, Z.: Matching pursuits with time-frequency dictionaries. IEEE Trans. Signal Process. 41(12), 3397–3415 (1993)
34. Cai, T.E., Wang, L.: Orthogonal matching pursuit for sparse signal recovery with noise. IEEE Trans. Inf. Theory 57(7), 4600–4688 (2011)
35. Donoho, D., Tsaig, Y.: Fast solution of coherence with noise. IEEE Trans. Inf. Theory 57(7), 4680–4688 (2011)
36. Melek, M., Khattab, A., Abu-Elyazed, M.F.: Fast matching pursuit for sparse representation-based face recognition. IET Image Process. 12(10), 1807–1814 (2018)
37. Deng, W., Hu, J., Guo, J.: In defense of sparsity based face recognition. In: 2013 IEEE Conference on Computer Vision and Pattern Recognition, pp. 399–406 (2013)
38. Zhang, Y., Peng, H.: One sample per person face recognition via sparse representation. IET Signal Processing 10, 1126–1134 (2016)
39. Iliadis, M., et al.: Sparse representation and least squares-based classification in face recognition. In: 22nd European Signal Processing Conference (EUSIPCO), pp. 526–530 (2014)
40. Fan, Z., et al.: Virtual dictionary based kernel sparse representation for face recognition. Pattern Recognit. 76, 1–13 (2018)
41. Cao, F., et al.: Pose and illumination variable face recognition via sparse representation and illumination dictionary. Knowl.-Based Syst. 107, 117–128 (2016)
42. Bilgazey, E., et al.: Improved face recognition using super-resolution. In: 2011 International Joint Conference on Biometrics (IJCB), pp. 1–7 (2011)
43. Uiboupin, T., et al.: Facial image super resolution using sparse representation for improving face recognition in surveillance monitoring. In: 24th Signal Processing and Communication Application Conference (SIU), pp. 437–440 (2016)
44. Yang, S., et al.: The CMU pose, sparse low-rank component-based representation for face recognition with low-quality images. IEEE Trans. Inf. Forensics Secur. 14, 251–261 (2019)
45. Cao, F., et al.: Fisher discrimination dictionary learning for sparse representation. In: International Conference on Image Processing, pp. 265–274 (2018)
46. Yadav, S., Vishwakarma, P.: Extended interval type-II and kernel based sparse representation method for face recognition. Expert Syst. Appl. 116, 265–274 (2018)