

Robust Face Recognition under Limited Training Sample Scenario using Linear Representation

Omer Iqbal¹, Waqas Jadoon¹, Zia ur Rehman¹, Fiaz Gul Khan¹, Babar Nazir¹,
Iftikhar Ahmed Khan¹

¹Department of Computer Science,
COMSATS University Islamabad,
Abbottabad Campus, 22060 - Pakistan
[E-mail: waqas_jadoon@ciit.net.pk]
*Corresponding author: Waqas Jadoon

*Received October 2, 2017; revised January 3, 2018; accepted January 26, 2018;
published July 31, 2018*

Abstract

Recently, several studies have shown that linear representation based approaches are very effective and efficient for image classification. One of these linear-representation-based approaches is the Collaborative representation (CR) method. The existing algorithms based on CR have two major problems that degrade their classification performance. First problem arises due to the limited number of available training samples. The large variations, caused by illumination and expression changes, among query and training samples leads to poor classification performance. Second problem occurs when an image is partially noised (contiguous occlusion), as some part of the given image become corrupt the classification performance also degrades. We aim to extend the collaborative representation framework under limited training samples face recognition problem. Our proposed solution will generate virtual samples and intra-class variations from training data to model the variations effectively between query and training samples. For robust classification, the image patches have been utilized to compute representation to address partial occlusion as it leads to more accurate classification results. The proposed method computes representation based on local regions in the images as opposed to CR, which computes representation based on global solution involving entire images. Furthermore, the proposed solution also integrates the locality structure into CR, using Euclidian distance between the query and training samples. Intuitively, if the query sample can be represented by selecting its nearest neighbours, lie on a same linear subspace then the resulting representation will be more discriminate and accurately classify the query sample. Hence our proposed framework model the limited sample face recognition problem into sufficient training samples problem using virtual samples and intra-class variations, generated from training samples that will result in improved classification accuracy as evident from experimental results. Moreover, it compute representation based on local image patches for robust classification and is expected to greatly increase the classification performance for face recognition task.

Keywords: Collaborative representation, robust classification, subspace learning, linear representation, face recognition

1. Introduction

Face recognition (FR) is one of the well-known applications of computer vision and is widely used in many areas especially for security such as secure valets, biometric identity recognition and many other electronic applications. Robust face recognition is considered challenging task due to two major problems. First problem arises due to the availability of limited or even a single training sample per class in the application of real-world face recognition such as national identity databases, law enforcement agencies etc. Therefore, because of insufficient training samples FR leads to poor classification [1][2][4][14][21]. Second problem, highlighted in the literature, is due to the corruption in face images such as occlusion, illumination and pose variations that makes face classification a challenging and complicated task [5][6][7][9]. The inferior performance of FR methods for single training sample per person (SSPP) problems is due to various reasons. Specifically, SSPP cannot give enough information about the query image for accurate modelling of variations between the test and training images [11][12][23][24][26]. Such variations are due to change in expressions and pose that exist in real-world scenarios. Due to these difficulties, traditional methods of FR are not effective under SSPP scenario [6][15][18][26]. Therefore, face image classification under limited number of samples and occluded faces conditions remains an open challenge in terms of computation time and classification accuracy.

Recently, representation based methods such as collaborative representation (CRC) [28] and sparse representation (SRC) [30] for face recognition have been popular among research community. These techniques can be generalized under a unified model of linear representation, which reconstruct a query image using a linear combination of available training samples. Subspace learning based approaches [38][40][41] for face recognition is another category of popular and vastly employed techniques among research community (such as PCA, LDA, LPP and LLE etc.). However, methods based on both of these categories required sufficient number of training samples from each class for effective classification, which unfortunality in case of real-world application is extremely difficult to meet. To address the small sample size problem of face recognition, Wan et al. [38] used fuzzy intrinsic and penalty graph to preserve nearest neighbor relationship among images to better characterize the compactness within a class and separate-ability between classes. Then maximum margin criterion (MMC) was used for feature extraction (low-dimensional space) and classification of face images. However, the technique did not explicitly address the SSPP problem and its performance is decreased significantly when the training samples per class are relatively small due to difficulties for preserving the topological structures in low-dimensional feature space. In [40], the authors proposed a joint learning framework for a compact and discriminative low feature space learning by integrating feature learning and image understanding for classification. However, performance of these methods suffers when applied to SSPP problem.

Another popular technique to address SSPP problem reported in literature is to use virtual or generic training sets. These techniques aims to generate extra samples from available training set for each class to extract the discriminatory information to model inter-class and intra-class variability among testing and training samples. For example, Jadoon et.al [6] have used optimally chosen variant basis generated from different class and training set by extending CR method. Deng et al. [26] models the variations between test and training images by utilizing an auxiliary intra-class variant dictionary by extending SRC classifier. However,

performance of these methods are limited to prior assumptions to extract discriminatory information that are too strong to assume in many cases [39]. Furthermore, these methods are based on holistic representation i.e. whole images are treated as high-dimensional vectors. Zhu et al. [44] proposed a multi-scale patched based CRC method by utilizing information obtained on various patch scales and then classification was done by combining recognition results of used patches. Recently various research studies have shown that local intrinsic representation can lead to robust and effective classification performance [18][39][42].

To address the above mentioned FR problems in the limited or single sample per person classification scenario, we propose to extend the Collaborative Representation (CR) approach [18] for image classification, which is a linear representation technique [3][10][22][28]. The reason of using a linear representation based method is due to the common observation that the test sample from single subject or unique class lie on a linear subspace of training sample to which it belongs [18][19][22][30][34]. The proposed method aims to compute representation based on local regions in the images as opposed to CR, which computes representation based on global solution involving entire images. Furthermore, the proposed solution also integrates the locality structure into CR, using Euclidian distance between the query and training samples. Intuitively, if the query sample is expressed by selecting its nearest neighbors, which lie on a same linear subspace, then the resulting representation will be more discriminate and accurately classify the query sample. We propose to integrate the symmetric images and virtual faces [11] [16] [32] in our CR framework for robust image classification. Our objective is to increase the number of training samples so that possible variations among testing samples can be modeled by training dictionary and the representation of test sample become more robust. To achieve this, we propose to generate virtual samples and model the testing samples variation using intra-class variations [8][26][33] among training samples (as the linear combination of existing samples) to overcome small sample size problem. It has been observed that symmetric face generation and linear combination of two sample solutions can perform robust image classification and can overcome illumination and pose variation problems [29][32]. To counter partial occlusion in facial images, we have partitioned the training and test images into ' M ' number of blocks (sub-images), through which each sub-image can take part in decision making of classification of test image. Thus, the partitions that are included in contiguous occlusion are represent by its corresponding training sub-image because that part does not belong to any samples presented in training data. Hence, noisy sub images do not participate in classification decision. Each partition image individually takes part in representation of a test sample and gives aggregate results. Combined with CR framework, the approach significantly results in higher accuracy than existing methods.

2. Related Work

To address the limited or single sample face recognition problem various techniques have been proposed in literature. The aim of these technique is to convert the specific (i.e., SSPP) problem into general face recognition problem by extending training set. These generalization techniques are briefly discussed below.

Jadoon et.al [6] have used extended Collaborative Representation (CR) to address the problem of single sample face recognition and proposed a novel approach in which a test image is represented linearly, by using all training samples and optimally chosen variant-bases generated from intra-variations from different classes and training dictionary. The variant

bases class is used to model possible facial variations among samples, such as illumination, disguise and facial expression while classification is performed by minimizing error reconstruction using l_2 -minimization. Yang Liu et.al [8] have used LGC (Learn-Generate-Classify) method to address STSPP (Single Training Sample per Person) problem. The technique picks up relationships among several samples of a single subject from a generic dictionary. After that, it calculates intra-class variation of the sample images present in gallery set with the help of the previously learned relationships. Using the predicted information, the synthetic images are produced. This converts the STSPP problem into multiple image samples scenario. Finally, the classification is done by applying SRC (Sparse Representation based Classification) method on the earlier generated multiple samples. Naseem et.al [16] have presented modified LRC (Linear Regression Classification) method for face identification problem. Authors have used the “single-object class lie on a linear subspace” concept to develop models that represent linear combination between test sample and class-specific images datasets. This model simply represented and measured the linearity by using l_2 -minimization method and decision of the specified test class is done by MRE (minimum reconstruction error). Another major problem faced in FR that is highlighted in the paper is “contiguous occlusion”. Occlusion makes the image corrupt, and parts of the image became noisy and unpredictable for proper classification. To address this issue, the author’s devised a modular LRC approach in which an image is divided into sub-blocks and each block is then treated as a sub-image. These sub-image are then individually used for classification decisions. This approach is dubbed as “Distance based evidence fusion” [16].

Zhang et.al [29], have proposed a method for generation of virtual faces to address the problem of single training sample per class. The method first produces symmetric faces by using symmetric transform followed by generation of virtual samples through linear combination of two samples. The extraction of features is done by using 2DPCA method and finally classification is based on the Nearest Neighbour (NN) method. Wright et.al [30] have used SRC (Sparse Representation based Classification) method to address face recognition problems, including occlusion. To some extent, the method handles errors i.e. to deal with test samples that are corrupt in the form of noise or disguise. Dealing with the sufficient feature extraction issue, sufficient number of training samples have been taken for a single subject to model the possible variations in query sample. An improvement has been made in the existing NNC (Nearest Neighbour Classifier) method by integrating it with the SRC method [31]. By integrating the SRC method, the classification accuracy significantly improved. The idea behind modification of NNC method is to complement correlation between training samples. Another improvement is that the distance between query sample and a training sample is estimated dependently. For representation of query sample, linear combination of every training sample is used in the improved method. Then the earlier calculated distance between query and training samples are used for the classification of test samples. The solution of linear equations is used for obtaining the coefficients of linear combination. The improved NNC method then calculates the distance of result (that came from multiplying coefficients with corresponding training sample) and the query sample, assuming that similar samples lie on a linear subspace. Finally the query sample are classified among the training sample that have them minimum distance between them.

Xu et.al [32] have worked on the limited number of training samples problem, that as explained earlier, is considered as one of the major challenges in face recognition tasks, especially in real world scenarios. To recover the variation gap between test and training face samples, the authors have used symmetric samples to increase the number of training samples for improved classification. The technique combines symmetric samples with original training

samples as a new training dictionary and classification is based on two-steps. The technique uses limited face image classes that are near or resemble more to the query sample for representation and use sparse representation based classification method for classification. This technique also utilizes SLF (Score Level Fusion) and FLF (Feature Level Fusion).

More recently, Liu et al. [39] have proposed the local structure based multi-phase collaborative representation classification (LS_MPCRC) to address the SSPP problem. Each image is divided into local blocks where each block is represented by overlapping patches. A query image patch is represented as the linear combination of training image patches using CRC from corresponding block, hence effectively generalizing the limited sample CRC problem to sufficient sample CRC problem. Finally, the classification confidence is improved by employing a multi-phase class selection scheme. Yang et al. [43] have extracted local adaptive convolution features by using convolution neural network (CNN) and use these discriminative features along with generic face dictionary by employing a joint and collaborative framework to address SSPP problem. A supervised autoencoder is stacked to get a deep neural network architecture [45] and used to extract the robust features for face recognition and representation. Face recognition is a one-shot learning task, it is not feasible to use the convolution neural networks such as VGG and GoogleNet, which have achieved state of the art results in the problem of object recognition where large samples per class are available. However, state of the art results in face recognition have been achieved by siamese neural networks [46][47]. These network consists of two CNNs with the same architecture (e.g. VGG without the final fully connected layers) and weights. An image pair is presented as input to the the network (one image per network) and the output generated by both networks is compared by using a distance measure. The cost function, called the triplet cost function, is high when the same images are classified as different or when different images as the same. The cost function is low when different images are classified as different or same images as the same. These neural networks learn the distance measure as the training progresses [48].

3. Representation Based Framework

3.1 Linear Representation

The linear representation of a query sample over the training samples dictionary can be expressed as follows:

Let y represent a vector $\in R^m$ of query sample for classification and belongs to any of the i^{th} class ($i = 1, 2, 3 \dots c$) in the training dictionary. Let matrix A represent the samples from the k^{th} training class having N_k number of training samples. Thus the samples from the k^{th} training class can be denoted as $A_k = [a_{k1} \dots a_{kN_k}] \in R^{m \times N_k}$. The training dictionary is formed by concatenating samples from all c classes into a matrix D as $D = [A_1, A_2, A_3, \dots, A_c]$. The query sample y can be represented as the linear combination of training samples from dictionary D .

$$y \approx Dx \quad (1)$$

$$y \approx \sum_{i=1}^c A_i x_i \quad (2)$$

Where $x_i \in R^N$ is the coefficient vector (also called coding vector) that tells about the contribution of training sample a_i (i^{th} image) in representing test image y . To calculate the

coefficient or coding vector, the problem can be formulated by minimize reconstruction errors between the query image and its gained representation [10] [28].

$$x = \operatorname{argmin} \|y - Dx\|_2 \quad (3)$$

$$x = (D^T D)^{-1} D^T y \quad (4)$$

The dictionary consists of c number of classes containing N number of total samples. The residuals against each i^{th} class calculated as the difference between the original and the estimated sample and by using the minimum residual the query sample y is classified to i^{th} training class.

$$r_i(y) = \|y - A_i x_i\|_2 \quad (5)$$

$$Id(y) = \operatorname{argmin}(r_i) \quad (6)$$

Where $i = 1, 2, 3 \dots c$. The CR (Collaborative Representation) framework [18][28] represents a query sample (y) collaboratively over the dictionary (D) where $D = [A_1, \dots, A_c]$ by using regularized l_2 -method as in Eqs. (7 and 8). The query sample y is classified to the i^{th} class if the residual error is minimum with respect to all classes using Eq (6).

$$x = \operatorname{argmin} \|y - Dx\|_2 + \lambda \|x\|_2 \quad (7)$$

$$x = (D^T D + \lambda I) D^T y \quad (8)$$

where λ is the small positive constant value and I represents the identity matrix.

Linear representation of a test sample over under-complete training samples dictionary gives unstable classification due to residuals having large values [18]. CR method can resolve small-sample size problem by collaboratively representing the test sample with all classes of training samples. The technique checks contribution of each class in representation and the class having least reconstruction error is showing similarity with the test sample [28].

If sample x and y are similar then the residual between x and y is very small, but a smaller amount of noise can make classification unstable (x and y become dissimilar because of noise that leads to misclassification). The CR approach can resolve this problem by adding regularization parameter on x to normalize the x and y samples. l_2 -regularization can outperform the state-of-the-art Compress Sensing methods based on l_1 -regularization [24][25][27], in term of computational complexity and classification accuracy [13][20][28].

3.2 Collaborative Neighbour Representation (CNR)

Face Recognition is typically a small-sample-size problem, each training class is under-complete [24][25]. Linear reconstruction of a query sample from a single class will lead to unstable classification due to large representational residual. Face images from different classes share certain similarities. Samples from other classes can be utilized in linear reconstruction of query sample to overcome small sample size problem and results in stable classification [28].

The Collaborative Neighbour Representation based classification approach [18] uses optimal neighbours from dictionary D against a query sample y for linear reconstruction of query sample. The bases are chosen automatically from the entire dictionary. The chosen base almost lies on a linear subspace on which the given query sample lies.

Mathematically the CNR problem can be defined as,

$$x = \operatorname{argmin} \left\{ \frac{1}{2} (\|y - Dx\|_2^2 + \sigma \sum_{i=1}^N x_i^2 \cdot \|y - a_i\|_2^2 + \lambda \|x\|_2^2) \right\} \quad (9)$$

The solution of above equation reconstruct the sample y from the neighbouring samples in dictionary D that span within linear subspace of sample y .

Whereas each training sample is denoted by a_i , the parameter ' λ ' and ' σ ' are small positive scalar values which are used as regularization parameters. The optimal solution of Eq. (9) can be derived by taking partial derivatives of (9) w.r.t. variable x and letting them equal to zero.

$$x = \left(D^T D + \lambda I + \sigma \begin{bmatrix} \|y - a_1\|_2 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \|y - a_N\|_2 \end{bmatrix} \right)^{-1} D^T y \quad (10)$$

Where $\begin{bmatrix} \|y - a_1\|_2 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \|y - a_N\|_2 \end{bmatrix} \in R^{N \times N}$ in eq. (10) is the matrix of diagonal non-zeros entries characterizes the distance between each sample in dictionary D and the query sample y .

Several studies have concluded that locality structure plays a crucial role in image classification. The test sample that is similar to the specific class of the training dictionary will lie on its linear subspace. Interpreting in another way, the similar class has the higher contribution in reconstruction of the test sample. This similarity is because of locality structure (nearest subspace), in Eq. (10) the similarity is based on Euclidian distance as distance of similar images is minimum. Including distance information in linear representation, enforces the higher contribution of similar classes into reconstruction (test and specific training class) as compare to other classes and hence results in lower reconstruction error with correct class during classification phase.

3.3 Proposed solution

The proposed solution is depicted in [Fig. 1](#). The main stages of the proposed solution consists of virtual samples generation, patch extraction, representation learning using CNR based classification and finally results aggregation.

3.3.1 Dealing Variations and Occlusion under limited sample scenario

Various studies have observed [\[5\]\[8\]\[9\]\[11\]\[12\]\[16\]\[21\]\[22\]](#) that increasing the number of training samples improves the classification accuracy, therefore, it is required to generate a sufficient number of training samples so that all possible variations of the test samples can be modelled by the training dictionary, thereby making the representation of test sample more vigorous. Under SSPP condition, generation of virtual samples from each class can overcome the variation problem between testing and training samples. Hence, intra-class variation can be determined by generating symmetric faces [\[32\]](#), and inter-class variation [\[6\]\[12\]](#) can be model by generating samples from linear combination of two samples. Patch based method is inducted into the collaborative representation framework to deal with face occlusion. Each patch (sub-image block) can independently take part in face image classification and finally the decision is made by aggregating the individual patch results.

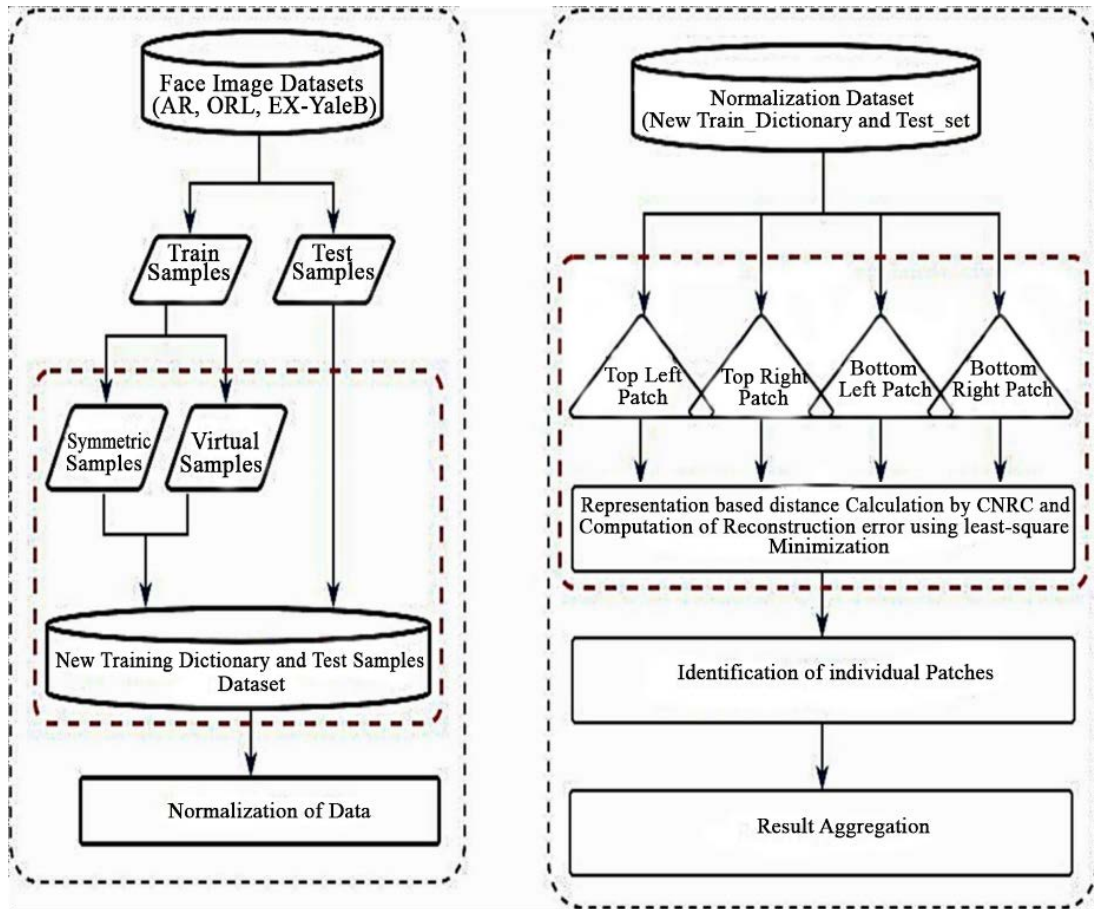


Fig. 1: Flow Chart of the Proposed Solution

3.3.2 Dealing Intra-Class Variation

The virtual samples can be generated by using symmetric face generation and linear combination of two samples from the original training set. To generate a symmetric face image from the original training sample, let a_i be the i^{th} sample in the training dictionary and let z_1 and z_2 be the symmetric faces of a_i , we have generated the two virtual faces by letting the left column of z_1 as the left half of a_i and the right column of z_1 is the mirror image of the left column of z_1 . Similarly, z_2 can be generated as the right column of z_2 is the original right half of a_i and the left column of z_2 is the mirror image of the right column of z_2 . The equation of symmetry is given as:

$$S(i, j) = R(i, V - j + 1) \quad (11)$$

Where U and V are number of rows and columns respectively and $i = 1 \dots U$ and $j = 1 \dots V$. the pixels location is denoted by $S(i, j)$. **Fig. 2(a)** shows some symmetric samples generated using eq. (11).



Fig. 2. (a) generation of symmetric samples from original samples

3.3.3 Handling Inter-Class Variation

The virtual samples generated from linear combination of two samples are used to increase the training samples. To deal with variations, such as pose and expression changes, the linear combination of samples from different classes can be utilized for robust classification [32]. Suppose that sample images a and b are taken from different classes, the linear fitting can be done by using Eq. (12):

$$z = \lambda a + (1 - \lambda)b, \text{ where } 0 \leq \lambda \leq 1 \quad (12)$$

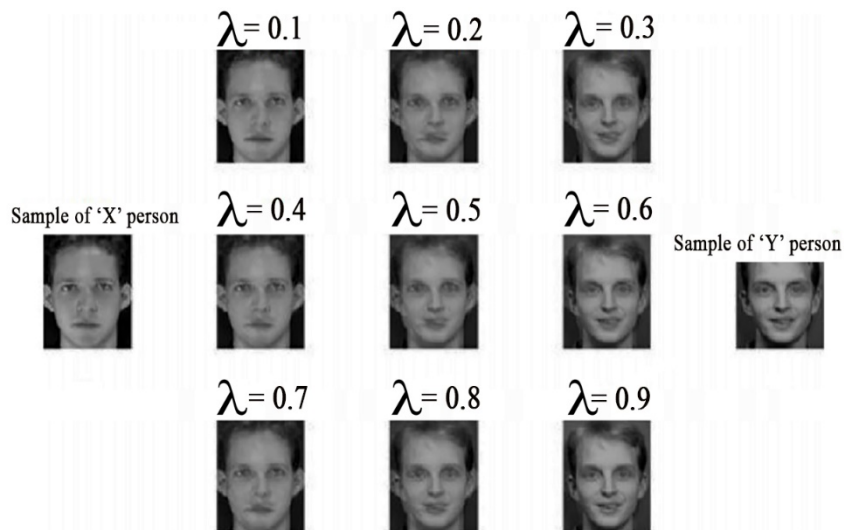


Fig. 2. (b) generation of virtual samples from linear combination of two samples

Sufficient number of representative training samples is generated to model the facial variations in images, by generating virtual samples, which leads to better classification results. Note that as shown in **Fig. 2 (b)**, not every linear fitting will produce real images. The images produced by the middle part of the linear fitting are apparently different from original images,

but the images on both ends of the linear fitting are very similar to original images, therefore, the selection of parameters is very important.

3.3.4 Robustness to Occlusion

We have enhanced the Collaborative Neighbour Representation (CNR) framework to effectively deal with challenges like contiguous occlusion and corruption. As explained earlier in the literature review [9][14], contiguous occlusion (partially noisy area of an image) in a sample image causes extreme deviation during comparison from correct classification of that sample, because a large portion of the image pixels is corrupted. Also, the amount of contaminated pixels is unknown. To address these issues, we present a solution, in which we integrate the patch based method [16] into CNR framework. The images are divided into multiple sub-images called as patches (as shown in Fig. 4) and each image sample is divided into four sub-images. Extracting patches from the single sample enables us to apply the CNR method independently on each sub image or patch. The method is applied individually on all the patches to compute the representation of test sample patches among all the training sample patches. So that the corrupted or noisy part will have no role in classification decision because of the large representational error whereas the clean patch (without corruption) will help in correct representation of the test sample. Integrating patch-based method with CNR framework has significantly improved the classification accuracy. The proposed approach is detailed in Algorithm 1:

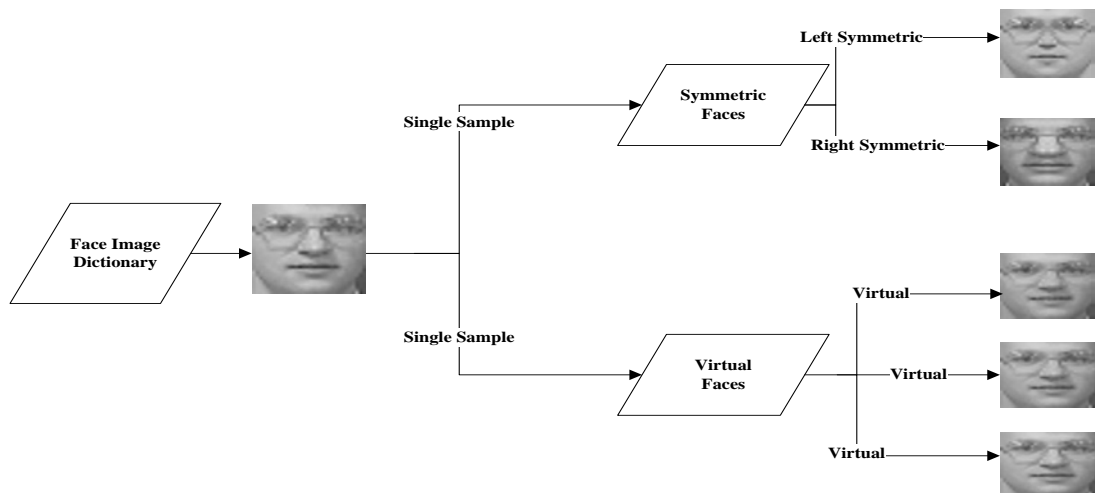


Fig. 3. Samples generation: Left & Right Symmetric and Virtual Samples

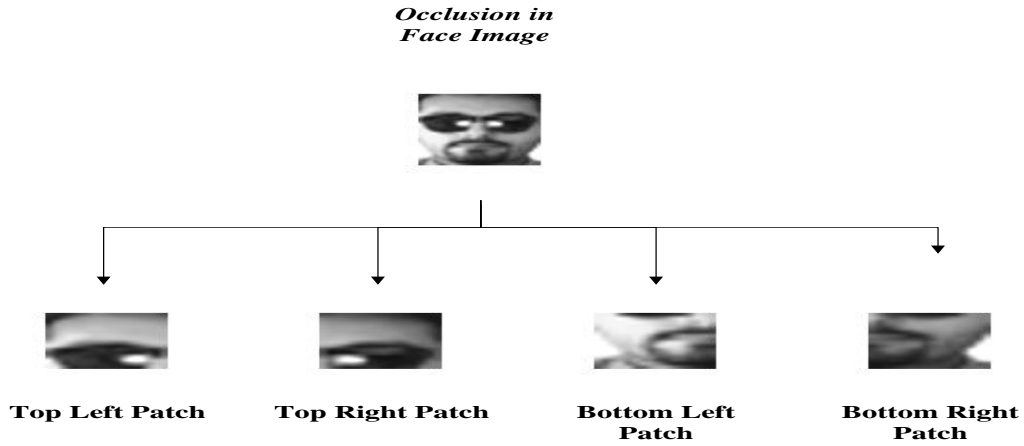


Fig. 4. Extraction of patches from sample Images

4. Dealing Insufficient and Occluded Samples

4.1 Methodology for Insufficient Training Samples

In order to deal with the single sample per person scenario, we have generated symmetric and virtual face image samples (shown in [Fig. 3](#)), so that a sufficient number of training samples per class exists in the training dictionary. The main purpose is to minimize the variations between the query and training samples, hence the representational error between query and the correct class is much smaller which results in stable classification.

The proposed CNR framework represents the query sample as the linear combination of all training samples. By integrating symmetrical samples along with original training samples, the variation gap between the query and training samples is reduced, contribution of correct training class in the reconstruction of query sample is increased, whereas the reconstruction error is minimized. All of these improvements lead to more accurate classification of query samples. Integration of virtual samples effectively handles expression variations among samples of same class and represents the query sample effectively by minimizing the within class deviation and maximization the between class deviation as shown ([Fig. 5](#) detailed the above approach) in the experimental results.

Algorithm 1: Proposed CNR Algorithm for Classification

1. **Input:** Query sample $y \in \mathbb{R}^m$ and face image samples as training dictionary $D = [A_1, A_2, A_3, \dots, A_C] \in \mathbb{R}^{m \times n}$ for C classes
2. Use symmetry transform for intra-class to generate symmetric image samples via Eq.(11)

$$Sym_{pk}(i, j) = A_k(i, V - j + 1) \{i = 1, \dots, U, j = 1, \dots, V\}$$

3. Use the linear combination method for inter-class to generate virtual training samples via Eq.(12)

$$Vir_k = \lambda a + (1 - \lambda) b, \quad 0 \leq \lambda \leq 1$$

4. Merge Original training, symmetric and virtual samples to form new training dictionary D
5. Normalize columns of dictionary D using z score method

$$z = (x - \mu) / \sigma$$

6. Extract patches from new training database for individual representation by using equation

$$I = face_{image} \left(1: \frac{m}{2}, 1: \frac{n}{2} \right)$$

7. Code query image y over D by: $x = (D^T D + \lambda I + \sigma D_{NH})^{-1} D^T y$

where $D_{NH} = \begin{bmatrix} \|y - a_1\|_2 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \|y - a_N\|_2 \end{bmatrix} \in \mathbb{R}^{N \times N}$ and N is the total number of samples in D

A diagonal matrix representing neighbourhood information. λ and σ are regularization parameters.

8. Compute the regularized reconstruction error:

$$r_i(y) = \|y - Dxi\|_2 / \|x\|_2$$

9. Aggregate result by using mode and sum rule functions.

10. **Output:** identify y to i^{th} class by:

$$Id(y) = \operatorname{argmin}_i [r_i(y)] \text{ for } i = 1, 2, \dots, C \text{ classes}$$

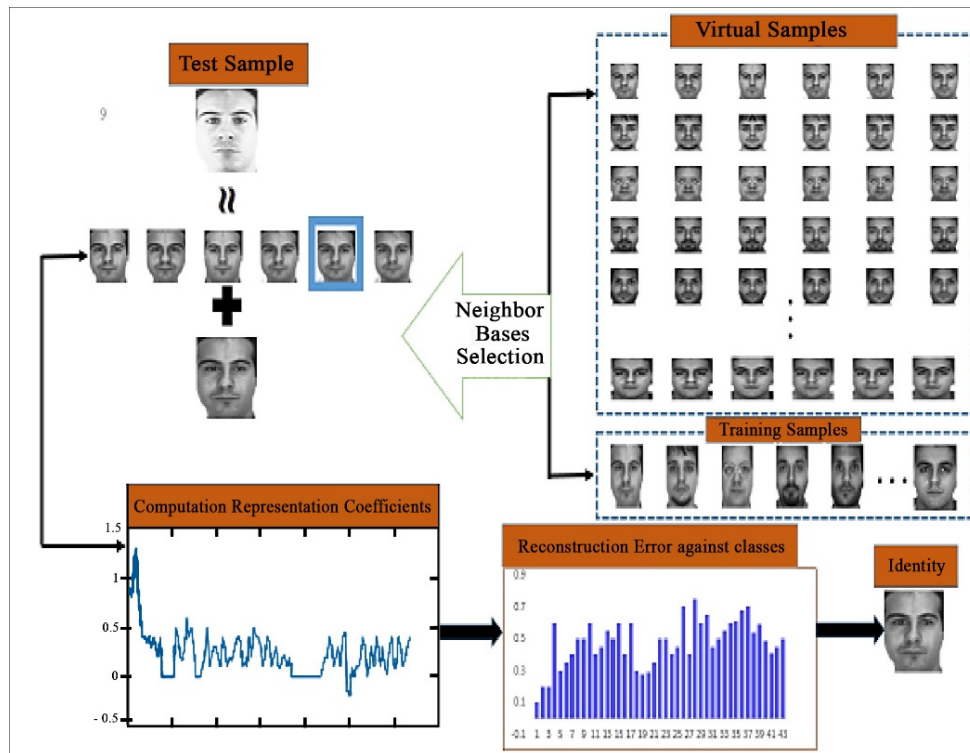


Fig. 5. Computation of Representation coefficients with Original + Virtual samples

4.2 Methodology for Occluded Samples

As explained earlier, classification of face images having occlusion is considered a challenging task. Contiguous occlusion or partially noised area refers to obstructed area in facial images through some objects such as glasses or scarves which makes it challenging to correctly classify the query image. To deal with this issue, we have devised a patch based solution by dividing each sample image into four blocks or patches (see Fig. 4) namely as top left patch, top right patch, bottom left patch and bottom right patch respectively. After integrating symmetric and virtual samples into single sample training dictionary, the corresponding patches are extracted from dictionary and query sample. Then applied CNR method individually on each path to compute the corresponding representation and get the classification result against each patch using minimum reconstruction error.

It has been observed during experimentation that patches without occlusion gave high representational coefficients, while the occluded parts, where the pixels were contaminated, have low representation coefficient as shown in Fig. 6. Finally, results are aggregated for final classification of the query sample.

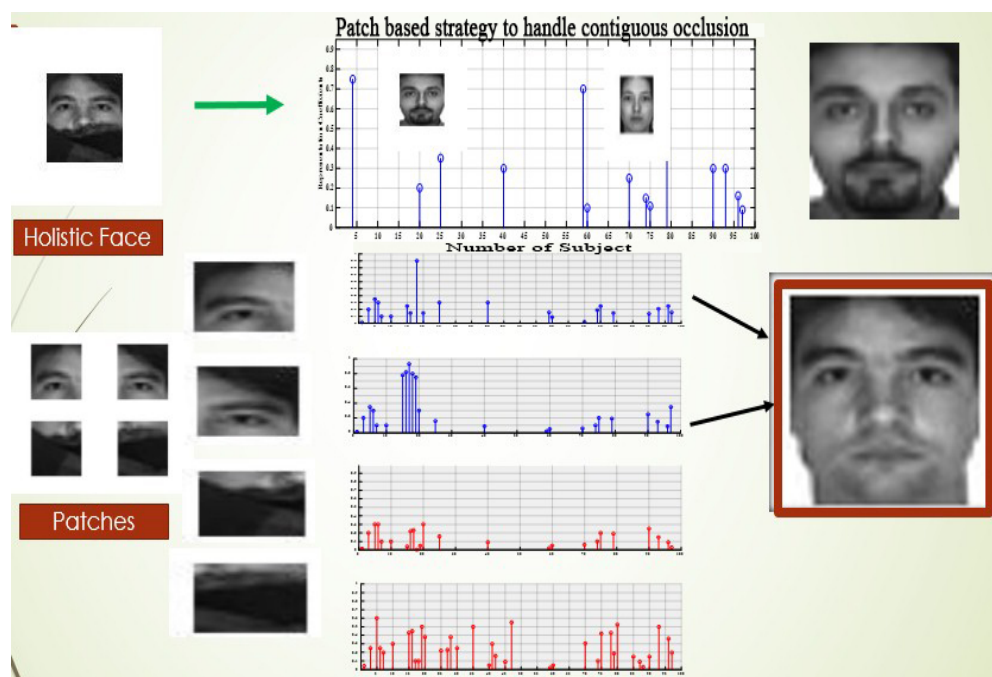


Fig. 6. Patch Based Solution and Impact of Individual representation

5. Experimental Results

5.1 Benchmark Datasets

This section discusses the detailed experimental results conducted on three publicly available, real world datasets of ORL [35], AR [36] and Extended-YaleB [37] to demonstrate the efficacy of the proposed method.

ORL Dataset consists of 400 images of 40 subjects, each subject have 10 images. The images include variations, such as facial expression, open or closed eyes, smiling or no smiling, facial details and pose variations. The training dictionary is built by choosing one neutral sample from each class, while remaining 360 samples are used as test samples. The symmetric and virtual samples are generated using training dictionary for the proposed scheme. Symmetric samples are used to handle the illumination effects while virtual samples deal with different kind of expression variations presented in query images. We have compared the proposed approach with some well-known existing methods namely Nearest Subspace (NS) [11], Collaborative Representation (CR) based Classification [28], 2-Dimensional Principle Component Analysis (2DPCA) [16][33] and a recently proposed method by Zhang et al. [29] to address SSPP face recognition problem by utilizing virtual samples. Table 1 shows the classification results of compared methods. The proposed CNR scheme achieved highest accuracy (73.9%) as compared to the Collaborative Representation (CR)(67%), Nearest Subspace (NS) (63%), 2-Dimensional Principle Component Analysis (2DPCA) (62%) and virtual faces [29] (68.22%) method.

AR Dataset consists of over 4000 images for 126 subjects, consisting of 65 men and 55 women taken in two different sessions with different facial expressions, lighting conditions and occlusions. We have used a subset of AR dataset containing 100 subjects (50 male and 50 female subjects). Each subject contains 13 images with variations. Therefore, the subset comprises of 1300 images in total. The single training sample per subject, having neutral frontal face image is utilized to build the training dictionary (a single neutral image from 100 subjects). Whereas for query samples, we made four subsets from remaining 1200 images that contain expression variations, illumination variations, occlusion and illumination and occlusion. Each sample image is cropped to size of 32x32 pixels. We have generated virtual samples using single sample per person training dictionary and integrated them into training dictionary.

Illumination subset

Table 1 shows the results of compared methods for illumination subset of AR dataset under single training sample per person (STSPP) face recognition scenario. As it can be seen that the proposed CNR method achieved highest classification accuracy (i.e., 92.5%) as compared with CR (82.4%), NS (78%), 2DPCA (50.2%) and Virtual faces[29](51%). Hence the results demonstrate that the proposed method effectively deals with illumination variations under STSPP face recognition problem.

Expression variation subset

Expression variation subset consists of 240 images that includes sad, happy, angry, screaming, smile and closed eye face variations. A single sample per person with neutral expression is utilized to construct the training dictionary.

Table 1 shows the classification results of compared methods on this subset. The proposed CNR approach achieved the highest accuracy of 88.3%, CR achieved 84.1%, NS 78.4% while 2DPCA and Virtual faces achieved 83% and 84%, respectively.

Holistic images Vs. Patched based classification

To demonstrate the efficacy of the proposed patch based classification scheme against complete images based classification scheme, we have used AR images subset that contains contiguous occlusion, partially noised and corrupt samples. **Table 1** shows the classification results of holistic vs. patched based approach of compared methods. The proposed CNR approach, that combines symmetric and virtual samples, achieved the highest accuracy as compared with other methods. However, the classification accuracy of patched based approach is superior to the holistic based approach. The main reason for this behaviour the large variation between query and training samples caused by occlusion in query images. When the holistic images are used, the methods based on linear representation are not able to accurately represent the query image due to large variations from correct class. However, patch based approach effectively deals with this issue by computing individual representation of each patch and only utilizing the clean patches for classification decision and discarding the occluded parts. Consequently, the aggregated classification results favour the correct class and improve classification results.

Extended YaleB Dataset includes 2414 images, each person has 64 images in the dataset and there are 38 individual subjects. Extended YaleB dataset particularly developed for different illumination effects. A variety of samples exist with a vast intensity range from very high to very low, and it becomes a challenging classification task for the existing methods specifically under single training sample per person scenario. In our experiments, we used resized samples of 32x32 pixels. Taking into account the real-world scenario, we have chosen

a single neutral image of each subject as a training sample (38 samples for training dictionary while remaining samples are used as testing samples). The classification results of proposed CRN, CR and 2DPCA are compared in **Table 2**. It can be observed from these results that the proposed method achieved the highest recognition rate i.e., 68% as compared to other methods for single training sample per subject scenario.

Table 1. Classification results of compared methods on ORL and AR datasets, bold value shows the highest classification achieved

| Dataset | CR | NS | 2DPCA | Virtual Faces[29] | CNR |
|--|-------|-------|-------|-------------------|--------------|
| ORL | 67% | 63% | 62% | 68.22% | 73.9% |
| AR | | | | | |
| Illumination subset | 82.4% | 78% | 50.2% | 51% | 92.5% |
| Expression subset | 84.1% | 78.4% | 83% | 84% | 88.3% |
| Occlusion subset | 78.3% | 68.2% | 21% | 27% | 80% |
| AR illumination and disguise subset | | | | | |
| Holistic images | 68% | 61% | - | - | 72% |
| Patched based | 76% | 67% | - | - | 85% |

We further experimented by increasing the number of training samples per subject between the ranges of two to five and remaining samples are used as test samples. As it can be seen from **Table 2**, that the proposed method performance is far better than the compared methods in all cases. The proposed method achieved 90% accuracy in case of 5-training sample per class, which is far better than the compared methods. The parameter values of λ and σ are chosen in the range of 0.001 to 0.1 through experimentation. Virtual faces [29] have used symmetric and virtual samples, by generating 9 virtual and 2 symmetric samples against each subject. Hence for large number of class, this results into overfitting of data and effects the classification accuracy. As more virtual samples are used for single class, it reduces the discriminant information between classes and results in misclassification of many quer samples. Furthermore, the method fails to effectively address the variations caused by illumination and occlusion as evident from the results shown in **Table 1** and **Table 2**.

Table 2. Classification accuracy on Extended YaleB dataset under different training sample per subject, bold value shows the highest classification achieved.

| Training Samples per Subject | CNR | CR | 2DPCA | Virtual Faces[29] |
|------------------------------|--------------|-------|-------|-------------------|
| Single | 68% | 62% | 32% | 39.5% |
| Two | 72.5% | 69.5% | 48% | 56% |
| Three | 78.5% | 75% | 62% | 65% |
| Four | 86% | 76% | 68% | 70% |
| Five | 90% | 79% | 73% | 77% |

5.2 Discussion

The experimental results confirm our assumption that by integrating virtual samples with the original training samples results in significant improvement in classification accuracy. Intuitively, the variation between the query and the training samples can be handled by virtual samples. Furthermore, the representation based on locality information (by inducting distance information) results in more accurate reconstruction of query sample from similar class bases which improves classification accuracy. Moreover, patch based method calculates the individual representation. Patches that achieved higher accuracy expressively helped in improving the aggregate classification result. The well-known techniques of CR, NS and 2DPCA have failed to effectively address the SSPP problem due to large variations among query and training samples. The virtual samples [29] specifically address the SSPP problem by generating virtual samples to extend the training set in order to cover the variations. However experimental results demonstrate that using large number of virtual samples do not result in increased classification performance. The classification rate of this method decreases significantly when the number of classes and query samples are increased. The method performance also suffers from variations caused by illuminations and occlusion. Compared to the virtual samples method, our method only utilizes three virtual samples that are very similar to the original image and two symmetric samples to cover illumination variations, hence effectively dealing with overfitting of data and resulting in significant improvement in classification as evident from experimental results. The average computational time of the proposed method against other techniques is shown in [Table 3](#).

Table 3. Average Computation time in seconds of CNR and comparative methods using whole Extended YaleB, ORL and AR face Datasets

| Dataset | Parameter | CNR | CR | NS | 2DPCA |
|----------------|------------|-------------|------|-------|-------|
| Extended YaleB | Time (sec) | 69.1 | 10.5 | 77 | 343.1 |
| | Speed Up | | | ~1.1 | ~4.7 |
| AR | Time (sec) | 64 | 15 | 77 | 180 |
| | Speed Up | | | ~1.2 | ~2.8 |
| ORL | Time (sec) | 9.8 | 1.4 | 11.5 | 50 |
| | Speed Up | | | ~1.17 | ~5.1 |

6. Conclusion

In this work, we proposed a method to address the robust face image classification under limited training sample using linear collaborative representation technique. To deal with variations, such as facial expressions, pose and illumination, among query and training samples, we used symmetric transformation and linear combination of samples to generate virtual samples. The integration of virtual samples with original samples models the limited training sample face recognition problem into sufficient samples problem. Thus the proposed solution effectively handles the intra-class and inter-class variations. For robust classification, the proposed patch based solution provides an efficient classification strategy. Moreover, to include local intrinsic structure into the computed representation, the proposed method incorporates locality information between query and training basis into the problem formulation, hence forcing similar basis with query sample for more contribution in reconstruction than dissimilar basis, which leads to more robust classification. Experimental results on real world

face datasets show that the proposed scheme is effective and outperforms some of the most widely used existing face classification methods.

For future direction, we will extend this work to address the problem of face recognition under SSPP scenario using deep learning. In recent literature, methods based on deep learning have demonstrated tremendous success for face recognition but these methods require large number of training samples per class to train the network. Thus how to use deep learning effectively for learning discriminant local features to address SSPP problem is one of our future direction. Another extension is to exploit the discriminant nature of specific facial landmark region, for example eyes and nose, that is more informative to human visual perception and have shown to have high detection rates under challenging facial variations.

References

- [1] X. Song, X. Yang, C. Shao and J. Yang, "Parity symmetrical collaborative representation-based classification for face recognition," *International Journal of Machine Learning and Cybernetics*, Volume 8, Issue 5, pp 1485–1492, October 2017. [Article \(CrossRef Link\)](#)
- [2] T. Larrain, J. S. Bernhard, D. Mery and K. W. Bowyer, "Face Recognition Using Sparse Fingerprint Classification Algorithm," *IEEE Transactions on Information Forensics and Security*, 1646 – 1657, 09 March 2017. [Article \(CrossRef Link\)](#)
- [3] H. Li, F. Shen, C. Shen, Y. Yang and Y. Gao, "Face recognition using linear representation ensembles," *Pattern Recognit.*, pp. 1–16, 2016. [Article \(CrossRef Link\)](#)
- [4] Y. Wen, W. Liu, M. Yang, Y. Fu, Y. Xiang and R. Hu, "Structured occlusion coding for robust face recognition," *Neurocomputing*, vol. 178, pp. 11–24, 2016. [Article \(CrossRef Link\)](#)
- [5] H. Zhang, F. Wang, Y. Chen, W. Zhang, K. Wang, and J. Liu, "Sample Pair based Sparse Representation Classification for Face Recognition," *Expert Systems with Applications*, vol. 45, pp. 352–358, 2016. [Article \(CrossRef Link\)](#)
- [6] W. Jadoon, L. Zhang, and Y. Zhang, "Extended collaborative neighbor representation for robust single-sample face recognition," *Neural Comput. Appl.*, vol. 26, no. 8, pp. 1991–2000, 2015. [Article \(CrossRef Link\)](#)
- [7] F. Dornaika and A. Bosaghzadeh, "Adaptive graph construction using data self-representativeness for pattern classification," *Information Sciences*, vol. 325, pp. 118–139, 2015. [Article \(CrossRef Link\)](#)
- [8] L. Yang, and I. Wassell, "A New Face Recognition Algorithm based on Dictionary Learning for a Single Training Sample per Person," in *Proc. of BMVC*, pp. 69–1, 2015. [Article \(CrossRef Link\)](#)
- [9] C. Ding and D. Tao, "Robust Face Recognition via multimodal deep face representation" *IEEE Transactions on Multimedia*, vol. 17, no. 11, pp. 2049–2058, 2015. [Article \(CrossRef Link\)](#)
- [10] S. Bektas, O. May, Y. Sisman and O. May, "The comparison of L1 and L2-norm minimization methods," *International Journal of the Physical Sciences*, Vol. 5(11), pp. 1721–1727, 2010. [Article \(CrossRef Link\)](#)
- [11] N. Zhu, T. Xu, L. Wei and T. Tang, "A nearest neighbor classifier based on virtual test samples for face recognition," *Opt. - Int. J. Light Electron Opt.*, vol. 126, no. 21, pp. 2799–2803, 2015. [Article \(CrossRef Link\)](#)
- [12] R. X. Ding, D. K. Du, Z. Huang, Z. Li and K. Shang, "Variational feature representation-based classification for face recognition with single sample per person," *J. Vis. Commun. Image Represent.* vol. 30, pp. 35–45, 2015. [Article \(CrossRef Link\)](#)
- [13] F. Dornaika, A. Bosaghzadeh, H. Salmane and Y. Ruichek, "Graph-based semi-supervised learning with Local Binary Patterns for holistic object categorization," *Expert Syst. Appl.*, vol. 41, no. 17, pp. 7744–7753, 2014. [Article \(CrossRef Link\)](#)
- [14] J. Wang, C. Lu, M. Wang and P. Li, "Robust Face Recognition via Adaptive Sparse Representation," *IEEE transactions on cybernetics*, vol. 44, no. 12, pp. 2368–2378, 2014. [Article \(CrossRef Link\)](#)

- [15] F. A. N. Liu, J. Tang, Y. A. N. Song, L. Zhang and Z. Tang, "Local Structure based Sparse Representation for Face Recognition," *ACM Transactions on Intelligent Systems and Technology (TIST)*, vol. 7, no. 1, pp.2, 2015. [Article \(CrossRef Link\)](#)
- [16] I. Naseem, R. Togneri and M. Bennamoun, "Linear regression for face recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 32, no. 11, pp. 2106–2112, 2010. [Article \(CrossRef Link\)](#)
- [17] A. Rahim, N. Hossain, T. Wahid, and S. Azam, "Face Recognition using Local Binary Patterns (LBP)," *Global Journal of Computer Science and Technology*, vol. 13, no. 4, pp. 469–481, 2013. [Article \(CrossRef Link\)](#)
- [18] J. Waqas, Z. Yi and L. Zhang, "Collaborative neighbor representation based classification using l(2)-minimization approach," *Pattern Recognition Letters*, vol. 34, no. 2, pp. 201–208, 2013. [Article \(CrossRef Link\)](#)
- [19] J. X. Mi and J. X. Liu, "Face recognition using sparse representation-based classification on k-nearest subspace," *PLoS One*, vol. 8, no. 3, p. e59430, 2013. [Article \(CrossRef Link\)](#)
- [20] Jian-Xun Mi, "Face image recognition via collaborative representation on selected training samples," *International Journal for Light and Electron Optics*, vol. 124, no. 18, pp. 3310–3313, 2013. [Article \(CrossRef Link\)](#)
- [21] Pengfei Zhu, Yang Meng, Lei Zhang and Il-Yong Lee, "Local generic representation for face recognition with single sample per person," in *Proc. of Asian Conference on Computer Vision*, pp. 34–50. Springer, Cham, 2014. [Article \(CrossRef Link\)](#)
- [22] J. Mi, D. Huang, B. Wang and X. Zhu, "The nearest-farthest subspace classification for face recognition," *Neurocomputing*, vol. 113, pp. 241–250, 2013. [Article \(CrossRef Link\)](#)
- [23] Y. Xu and Q. Zhu, "A simple and fast representation-based face recognition method," *Neural Computing and Applications*, pp. 1543–1549, 2013. [Article \(CrossRef Link\)](#)
- [24] A. Y. Yang, J. Wright, S. Member, Y. Ma, S. Member, and S. S. Sastry, "Feature Selection in Face Recognition : A Sparse Representation Perspective," *IEEE Trans. Pattern Anal. Mach. Intell.*, no. August, pp. 1–34, 2007. [Article \(CrossRef Link\)](#)
- [25] A. Y. Yang, Z. Zhou, A. G. Balasubramanian, S. S. Sastry and Y. Ma, "Fast L1-Minimization Algorithms for Robust Face Recognition," *IEEE Transactions on Image Processing*, vol. 22(8), pp. 3234–3246, 2013. [Article \(CrossRef Link\)](#)
- [26] W. Deng, J. Hu and J. Guo, "Extended SRC : Undersampled Face Recognition via Intra-class Variant Dictionary," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 34, no. 9, pp. 1864–1870, 2012. [Article \(CrossRef Link\)](#)
- [27] E. Bilgazyev, B. Efraty, S. Shah and I. Kakadiaris, "Sparse Representation-Based Super Resolution for Face Recognition At a Distance," *Proceedings Br. Mach. Vis. Conf.* 2011, pp. 52.1–52.11, 2011. [Article \(CrossRef Link\)](#)
- [28] L. Zhang, M. Yang and X. Feng, "Sparse Representation or Collaborative Representation : Which Helps Face Recognition ?," in *Proc. of IEEE international conference on Computer vision (ICCV)*, pp. 471–478, 2011. [Article \(CrossRef Link\)](#)
- [29] T. Zhang, X. Li and R. Guo, "Producing virtual face images for single sample face recognition," *Opt. - Int. J. Light Electron Opt.*, vol. 125, no. 17, pp. 5017–5024, 2014. [Article \(CrossRef Link\)](#)
- [30] J. Wright, A. Y. Yang, A. Ganesh, S. S. Sastry and Y. Ma, "Robust face recognition via sparse representation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 31, no. 2, pp. 210–227, 2009. [Article \(CrossRef Link\)](#)
- [31] Y. Xu, Q. Zhu, Y. Chen and J.S. Pan, "An improvement to the nearest neighbor classifier and face recognition experiments," *Int. J. Innov. Comput. Inf. Control*, vol. 9, no. 2, pp. 543–554, 2013. [Article \(CrossRef Link\)](#)
- [32] Y. Xu, X. Zhu, Z. Li, G. Liu, Y. Lu and H. Liu, "Using the original and ' symmetrical face ' training samples to perform representation based two-step face recognition," *Pattern Recognition*, vol. 46, no. 4, pp. 1151–1158, 2013. [Article \(CrossRef Link\)](#)
- [33] Nhat Vo Dinh Minh and SungYoung Lee, "Kernel-based 2DPCA for face recognition," in *Proc. of IEEE International Symposium on Signal Processing and Information Technology*, pp. 35–39, 2007. [Article \(CrossRef Link\)](#)

- [34] F. S. Samaria, & A. C. Harter, "Parameterization of a stochastic model for human face identification," in *Proc. of Proceedings of the Second IEEE Workshop on Applications of Computer Vision*, pp 138–142, 1994. [Article \(CrossRef Link\)](#)
- [35] AT&T LAB of the University of Cambridge, UK, ORL face database [EB/OL], 2010. [Article \(CrossRef Link\)](#)
- [36] A. M. Martinez and R. Benavente, "The AR Face Database," CVC Technical Report 24, 1998. [Article \(CrossRef Link\)](#)
- [37] A. Georghiades, P. Belhumeur and D. Kriegman, "From Few to Many: Illumination Cone Models for Face Recognition under Variable Lighting and Pose," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 23, no. 6, pp. 643-660, June 2001. [Article \(CrossRef Link\)](#)
- [38] M. Wan, Z. Lai, G. Yang, Z. Yang, F. Zhang and H. Zheng, "Local graph embedding based on maximum margin criterion via fuzzy set," *Fuzzy Sets and Systems*, vol. 318, pp. 120-131, 2017. [Article \(CrossRef Link\)](#)
- [39] F. Liu, J. Tang, Y. Song, Y. Bi, and S. Yang, "Local structure based multi-phase collaborative representation for face recognition with single sample per person," *Information Sciences*, vol. 346, pp. 198-215, 2016. [Article \(CrossRef Link\)](#)
- [40] Z. Li, J. Liu, J. Tang, and H. Lu, "Robust structured subspace learning for data representation," *IEEE transactions on pattern analysis and machine intelligence*, vol. 37, no. 10, pp. 2085-2098, 2015. [Article \(CrossRef Link\)](#)
- [41] Z. Li, J. Liu, Y. Yang, X. Zhou and H. Lu, "Clustering-guided sparse structural learning for unsupervised feature selection," *IEEE Transactions on Knowledge and Data Engineering*, vol. 26(9), pp. 2138-2150, 2014. [Article \(CrossRef Link\)](#)
- [42] Z. Zhengjuan and W. Jadoon, "Intrinsic Structure based Feature Transform for image classification," *Journal of Visual Communication and Image Representation*, Vol. 38. pp. 735–744, 2016. [Article \(CrossRef Link\)](#)
- [43] M. Yang, X. Wang, G. Zeng and L. Shen, "Joint and collaborative representation with local adaptive convolution feature for face recognition with single sample per person," *Pattern Recognition*, Vol. 66. pp. 117-128, 2017. [Article \(CrossRef Link\)](#)
- [44] P. Zhu, L. Zhang, Q. Hu and S. C. Shiu, "Multi-scale patch based collaborative representation for face recognition with margin distribution optimization", in *Proc. of European Conference on Computer Vision*, Springer, Berlin, Heidelberg. pp. 822-835, 2012. [Article \(CrossRef Link\)](#)
- [45] S. Gao, Y. Zhang, K. Jia and Y. Zhang, "Single sample face recognition via learning deep supervised autoencoders," *IEEE Transactions on Information Forensics and Security*, vol. 10, no. 10, pp. 2108-2118, 2015. [Article \(CrossRef Link\)](#)
- [46] G. Koch, Z. Richard and S. Ruslan, "Siamese neural networks for one-shot image recognition," in *Proc. of ICML Deep Learning Workshop*, Vol. 2. 2015. [Article \(CrossRef Link\)](#)
- [47] Y. Taigman, M. Yang, M.A. Ranzato and L. Wolf, "Deepface: Closing the gap to human-level performance in face verification," in *Proc. of Proceedings of the IEEE conference on computer vision and pattern recognition*, pp.1701-1708, 2014. [Article \(CrossRef Link\)](#)
- [48] F. Schroff, K. Dmitry and P. James, "Facenet: A unified embedding for face recognition and clustering," in *Proc. of Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2015. [Article \(CrossRef Link\)](#)



Omer Iqbal is a graduate student at Department of Computer Science, COMSATS University, Pakistan. His research interest includes image processing and machine learning.



Dr. Waqas Jadoon received the Ph.D. degree in Computer Science from Sichuan University China in 2014. Currently, he is an Assistant Professor at COMSATS Institute of Information Technology, Abbottabad, Pakistan and is an approved Ph.D Supervisor from Higher Education Commission (HEC) Pakistan. His research interests focus on Pattern Recognition, Image Processing, theory and applications of Machine Learning.



Dr. Zia ur rehman is an assistant professor at the Department of Computer Science, COMSATS Institute of Information Technology, Abbottabad, Pakistan. He did his PhD in information systems from Curtin University, Perth, Australia. He also holds a Master of Computer Science and an MSc. In Applied Mathematics from the University of Balochistan, Quetta, Pakistan.



Dr. Fiaz Gul did his graduation and MS from COMSATS Institute of Information Technology Abbottabad in the field of Computer Science. For specialization master and Doctorate he won the Higher Education Commission (HEC) scholarship under the project UESTP for Politecnico di Torino Italy. Currently he is serving as an Assistant Professor in Computer Science Department at COMSATS Abbottabad, Pakistan.



Dr. Babar Nazir is Assistant Professor in Computer Science Department at the COMSATS Institute of Information Technology, Abbottabad, Pakistan. He obtained his Ph. D in Computer Science from the Universiti Teknologi Petronas, Malaysia in 2011. He received his M.S in Computer Science from the COMSATS Institute of Information Technology, Abbottabad, Pakistan in 2007. His research interest includes: resource management and job scheduling in Cloud Computing, Grid Computing, and Cluster Computing, Communication protocols for Wireless Sensor networks, Communication protocols for Mobile Adhoc Networks (Bluetooth, VANETs). He has published more than 50 papers in reputed international conferences and journals.



Dr. Iftikhar Ahmed Khan did his PhD in the field of 'Human Computer Interaction' from Brunel University West London UK in 2009. Dr. Khan did MCS in the field of computer sciences from COMSATS Abbottabad in 2004. His areas of research are Human Computer Interaction, Affective Computing, Image processing and, Software Engineering. He has published paper in various reputed international conferences and journals. He is currently working as an ASSISTANT PROFESSOR in the Department of Computer Sciences, COMSATS Institute of Information Technology Abbottabad. He worked as a SOFTWARE ENGINEER as well for a year in COMSATS Information Technology Centre.