Robust Label Consistent Dictionary Learning for Face Recognition

Sowjanya Modalavalasa, Upendra Kumar Sahoo, Ajit Kumar Sahoo
Electronics and Communication Department
National Institute of Technology
Rourkela, Odisha
India, 769008
Email: sowjanya.modalavalasa@gmail.com,(sahooupen,ajitsahoo@nitrkl.ac.in)

***Abstract*—The Label Consistent K-Singular Value Decomposition (LC-KSVD) algorithm, which has been introduced recently has shown better results for learning a discriminative dictionary for recognition and classification by considering the label consistency constraint in the cost function. However this approach assumes Guassian distribution for the coding residual, which might not be true in the practical Face Recognition system due to lighting, occlusion/disguise and expression variations. In this paper, we propose a new Robust Label Consistent Dictionary Learning (RLC-DL) algorithm, which assumes that the coding residual and coefficients are identically distributed, independent and tries to find a Maximum Likelihood Estimation of the coding problem. The proposed algorithm is evaluated on the publicly available face datasets and it shows superior performance than state of the art algorithms available including LC-KSVD for face recognition.**

***Index Terms*—Face Recognition, Robust Label Consistent Dictionary, Dictionary Learning**

I. Introduction

Face recognition is one of the most active areas in computer vision and machine learning due to its immense range of applications ranging from Access Control, General identity verification, Surveillance and so on. The idea of face recognition is challenging due to disguise, occlusion, lighting, intensity variations and bad quality of face images. Sparse coding and dictionary learning have been successively employed in the face recognition problem. The key idea of face recognition using dictionary learning [1] [2] is to project the high dimensional face data [3] onto low dimensional sparse subspace and explore its discriminative characteristics for representation. For classification, the test face has to be projected on to that subspace followed by any classifier. The Sparse Representation based Classification (SRC) by Wright *et al.* [4] has shown promising results for face recognition. In this SRC method, the dictionary is obtained by concatenating the training samples from all the subjects available and the test image is represented as the sparse linear combination of those dictionary atoms by choosing the best reconstruction error. The dictionary could be very big in most practical cases which increases the coding complexity and during noisy conditions, the query image may not be well represented. The dictionary in SRC is huge and is fixed, whereas many adaptive dictionary learning algorithms [5] [6] have been proposed in the literature.

The K-SVD algorithm proposed by Aharon *et al.* [7] has shown promising results for signal representation, image denoising, image compression [2] [5] [8] and so on. Since K-SVD minimizes the reconstruction error, it is only capable of best representation of the signals but not preferable for classification. Zhang *et al.* [9] included the classification error to the cost function, which is in turn optimized by using K-SVD and it gave promising results for face recognition. Another similar extension by Jiang *et al.* [10] made the dictionary and sparse coefficients to be discriminative due to the label consistent constraint in the optimization function. A structured dictionary considering the Fisher discrimination constraint has been introduced by Yang *et al.* [6], which diminishes the intra class variations and maximizes the inter class variations [11] of the coding coefficients. All the methods discussed till now assume that the fiedility error follows either Gaussian or Laplacian distribution, which will fail in practical face representation systems due to disguise or occlusion, noise, illumination/expression/pose variations. Yang *et al.* [12] generalized the coding error probability density function and obtained the maximum likelihood estimation(MLE) of the sparse coding problem for robust face recognition. Undersampled face recognition has been addressed by Deng *et al.* [13], Wei *et al.* [14] by introducing an intraclass variant dictionary and Robust Auxiliary Dictionary Learning respectively.

The label consistent K-SVD [10] will force the sparse codes and dictionary to be discriminant, which makes this algorithm to be much preferable for classification. The unified objective function of LC-KSVD assumes its coding error to be Gaussian, which might not be true in all scenarios. We proposed Robust Label Consistent Dictionary Learning (RLC-DL), which generalized the coding error distribution function and achieved better accuracy compared to LC-KSVD for face recognition. The notations considered in this work are, bold capital letters for representing matrices, bold small letters for representing vectors, and samll letters for scalars.

This paper is organized as follows. The problem description has been provided in sec II. Section III explains the proposed work. In Section IV, we have presented the experimental results on two publicly available datasets and compared them with many face recognition algorithms methods. Section V concludes the paper.

II. Problem Description

A. Label Consistent K-Singular Value Decomposition

Label Consistent K-SVD [10] is an ehanced version of Discriminative K-SVD, where a classification error is included in cost function along with reconstruction error. The optimization function of LC-KSVD contains classification error, discriminative sparse code error and reconstruction error [10]. The presence of discriminative sparse code error forces the sparse codes of same classes to be similar and sparse codes of different classes to be distinct, which in-turn results in a discriminant dictionary. Let $\mathbf{Y}$ denote the set of N input signals each of dimension n, i.e. $\mathbf{Y} = [\mathbf{y}_1, \mathbf{y}_2 .....\mathbf{y}_N] \in R^{n\times N}$. Let $\mathbf{D}$ be the dictionary with K atoms, i.e. $\mathbf{D} = [\mathbf{d}_1, \mathbf{d}_2 \; ..... \; \mathbf{d}_K] \; \in R^{n\times K}$. The dictionary is overcomplete ($K > n$) and sparsity constraint is imposed to get unique solution out of the infinite possible solutions. The sparse codes of the input $\mathbf{Y}$ are given by $\mathbf{X} = [\mathbf{x}_1, \mathbf{x}_2 .....\mathbf{x}_N] \in R^{K\times N}$, and L is the sparsity constraint factor.
The cost function of LC-KSVD1 [10] is given as:

$$\begin{aligned}\langle \mathbf{X}, \mathbf{D}, \mathbf{A}\rangle = \underset{\mathbf{X},\mathbf{D},\mathbf{A}}{\arg\min} \; \|\mathbf{Y} - \mathbf{D}\mathbf{X}\|_2^2 + \theta \|\mathbf{Q} - \mathbf{A}\mathbf{X}\|_2^2 , \\ s.t. \forall i, \|\mathbf{x}_i\|_0 \le L\end{aligned} \tag{1}$$

The cost function of LC-KSVD2 [10] is given as:

$$\begin{aligned}\langle \mathbf{X}, \mathbf{D}, \mathbf{A}, \mathbf{W}\rangle = \underset{\mathbf{X},\mathbf{D},\mathbf{A},\mathbf{W}}{\arg\min} \; \|\mathbf{Y} - \mathbf{D}\mathbf{X}\|_2^2 + \theta \|\mathbf{Q} - \mathbf{A}\mathbf{X}\|_2^2 \\ + \gamma \|\mathbf{H} - \mathbf{W}\mathbf{X}\|_2^2 , \;\; s.t. \forall i, \|\mathbf{x}_i\|_0 \le L\end{aligned} \tag{2}$$

Where the first error term indicates the reconstruction error [7], second error term signifies the discriminative sparse code error [10] and finally the last one indicates the classification error [9]. θ and γ control the contribution of discriminative sparse code error and classification error. $\mathbf{A}$ is the linear transformation matrix and $\mathbf{H}$ is the label matrix with the label details of each class in $\mathbf{Y}$. $\mathbf{Q}$ is the discriminative sparse code matrix given as $\mathbf{Q} = [\mathbf{q}_1, \mathbf{q}_2 .....\mathbf{q}_N]$. $\mathbf{q}_i$ will have non zero indices where $\mathbf{d}_i$ and $\mathbf{y}_i$ have same label. If $\mathbf{y}_1$, $\mathbf{y}_2$ and $\mathbf{d}_1$, $\mathbf{d}_2$ are from first class and $\mathbf{y}_3$, $\mathbf{y}_4$ and $\mathbf{d}_3$, $\mathbf{d}_4$ are from class 2, then $\mathbf{Q}$ will become:

$$\mathbf{Q} = \begin{bmatrix} 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 1 & 1 \end{bmatrix}$$

The unified objective function is defined by reformulating the equation (2) as

$$\begin{aligned}\langle \mathbf{D}_{new}, \mathbf{X}\rangle = \underset{\mathbf{D}_{new},\mathbf{X}}{\arg\min} \left\{ \|\mathbf{Y}_{new} - \mathbf{D}_{new}\mathbf{X}\|_2^2 \right\} , \\ s.t. \forall i, \|\mathbf{x}_i\|_0 \le L\end{aligned} \tag{3}$$

where $\mathbf{Y_{new}} = \left[\mathbf{Y}; \sqrt{\theta}\mathbf{Q}; \sqrt{\gamma}\mathbf{H}\right], \mathbf{D_{new}} = \left[\mathbf{D}; \sqrt{\theta}\mathbf{A}; \sqrt{\gamma}\mathbf{W}\right]$. The K-SVD algorithm is applied to solve the above problem. Now the matrices $\mathbf{D}, \mathbf{A}, \mathbf{W}$ are denormalized and classification is done using the classifier parameter $\mathbf{W}$.

$$label = \mathbf{W} \times \hat{\mathbf{X}}_i$$

B. Limitation of LC-KSVD

As mentioned in equations (1) and (2), LC-KSVD considers that the coding residual error follows Gaussian distribution and utilizes MLE to obtain the solution. Most of the classical approaches follow either l_1-norm or l_2-norm of the fidelity error in the optimization problem. When the coding error is assumed to follow Gaussian distribution, the maximum likelihood estimation of the fidelity term is represented by the l_2-norm. If the coding error is assumed to follow Laplacian distribution, then the maximum likelihood estimation of the fidelity term is represented by the l_1-norm. In practical scenario of the face recognition system, the error may or may not follow either the Gaussian or the Laplacian distribution due to occlusion, disguise, noise, variations in pose, expression and illumination. To enhance the performance of LC-KSVD in such scenarios, a more generalized maximum likelihood estimation of the residual error is needed.

III. Proposed Framework

To enhance the performance of the LC-KSVD algorithm in above mentioned scenario, we propose a Robust Label Consistent Dictionary Learning by finding the MLE solution of coding coefficients. The work has been inspired by [12] [14], where the maximum likelihood estimation solution of the coding residual in the cost function is solved by iteratively re-weighted sparse coding and dictionary learning techniques. The optimization function can be formulated as below:

$$\langle \mathbf{X}, \mathbf{D}, \mathbf{A}\rangle = \underset{\mathbf{X},\mathbf{D},\mathbf{A}}{\arg\min} \; \rho\left(\mathbf{Y} - \mathbf{D}\mathbf{X}\right) + \rho\left(\mathbf{Q} - \mathbf{A}\mathbf{X}\right) + \lambda \|\mathbf{X}\|_1 \tag{4}$$

where the first term indicate the residual function of reconstruction error and the second term signifies the residual function of discriminative sparse code error. The last term in the above formulation is the l_1-regularization term of the sparse code error.

A. Optimization

Let us consider that the above two coding error distributions can be combined to form an unknown probability distribution and find its maximum likelihood estimation solution. The two residual functions can be denoted in a unified way as shown below

$$\underset{\mathbf{X},\mathbf{D}^{new}}{\arg\min} \; \rho\left(\mathbf{Y}^{new} - \mathbf{D}^{new}\mathbf{X}\right) \; + \lambda \|\mathbf{X}\|_1 \tag{5}$$

Where the residual error defined as $\mathbf{e} = \mathbf{Y}^{new} - \mathbf{D}^{new}\mathbf{X}$ and

$$\mathbf{Y}^{new} = \begin{pmatrix} \mathbf{Y} \\ \mathbf{Q} \end{pmatrix} \quad and \;\; \mathbf{D}^{new} = \begin{pmatrix} \mathbf{D} \\ \mathbf{A} \end{pmatrix}$$

Here $\mathbf{Q}$ is the discriminative sparse code matrix and $\mathbf{A}$ is a linear transformation matrix. The residual function is defined as

$$\rho(e) = \sum_{k=1}^{n} \rho(e_k)$$

After solving the above equation using Maximum likelihood estimation as in [13] [15], equation (5) can be defined as

$$\langle \mathbf{D}^{new}, \mathbf{X} \rangle = \underset{\mathbf{D}^{new}, \mathbf{X}}{\arg\min} \left\| \mathbf{W}^{1/2} \left(\mathbf{Y}^{new} - \mathbf{D}^{new}\mathbf{X} \right) \right\|_2^2 + \lambda \|\mathbf{X}\|_1 \tag{6}$$

where $\mathbf{e}_i = \mathbf{y}_i^{new} - \mathbf{D}^{new}\mathbf{x}_i$ and the diagonal matrix $\mathbf{W}$ is given as

$$\mathbf{W} = diag\left\{ w(e_1), w(e_2), w(e_3), ..., w(e_n) \right\} \tag{7}$$

A variety of residual functions are available in robust M-estimatiors theory [16]. In our work we have considered the logistic loss function and the huber loss function for robust estimation.

1) Logistic Loss: The residual function [4] considered is

$$\rho(e_k) = -\frac{1}{2\mu} \left(\ln\left(1 + \exp\left(-\mu e_k^2 + \mu\delta\right)\right) - \ln\left(1 + \exp \mu\delta\right) \right) \tag{8}$$

Now the weight function calculated using the equality $w(e_k) = \frac{d\rho_c(e_k)}{de_k}\frac{1}{e_k}$, is defined as:

$$w(e_k) = \frac{\exp\left(\mu\left(\delta - e_k^2\right)\right)}{1 + \exp\left(\mu\left(\delta - e_k^2\right)\right)} \tag{9}$$

The constants μ and δ are chosen in such a way that less weightage is assigned to ouliers. As $e_k \to 0$ then $w(e_k) \to 1$, as it suggests the presense of inliers. For larger valuers of the error(outliers) the weight should be very small. To satisfy theses criterions, the standard setting in the literature [4] [12] is followed. Let $\mu\delta = C_{\mu\delta}$ and $C_{\mu\delta}$ is chosen to be greater than or equal to 8 in all cases. The squared error vector is sorted $e_s = sort\left(e_1^2, e_2^2, ..., e_n^2\right)$ and for $\tau \in [0.6, 0.8]$, $l = length(e_s)$ and $i = ceil(\tau l)$, δ is defined as the largest i_{th} integer of e_s.

2) Huber loss: The huber function is defined as

$$\rho_c(e_k) = \begin{cases} e_k^2 & if \quad |e_k| \le c \\ 2c|e_k| - c^2 & if \quad |e_k| > c \end{cases} \tag{10}$$

Now the weight function of huber loss obtained as $w(e_k) = \frac{d\rho_c(e_k)}{de_k}\frac{1}{e_k}$

$$w(e_k) = \begin{cases} 1 & if \quad |e_k| \le c \\ \dfrac{2c\,\mathrm{sgn}(e_k)}{e_k} & if \quad |e_k| > c \end{cases} \tag{11}$$

The huber function [16] is qudratic in the central region and is linearly related in the other regions. So it gives less weightage to the outliers and more weightage to the inliers, resulting in a robust estimation. The constant c is chose empherically as per the requirement.

B. Sparse coding

To compute the sparse codes, fix the dictionary $\mathbf{D}$ and optimize the cost function in equation (5) with respect to the sparse code vector $\mathbf{X}$. So the formulation is equivalent to solving the problem mentioned below:

$$\langle \mathbf{x}_i \rangle = \underset{\mathbf{x}_i}{\arg\min}\, \rho\left(\mathbf{y}_i^{new} - \mathbf{D}^{new}\mathbf{x}_i\right) + \lambda\|\mathbf{x}_i\|_1$$

The maximum likelihood solution of the above equation is given as

$$\langle \mathbf{x}_i \rangle = \underset{\mathbf{x}_i}{\arg\min} \left\| \mathbf{W}^{1/2}\left(\mathbf{y}_i^{new} - \mathbf{D}^{new}\mathbf{x}_i\right) \right\|_2^2 + \lambda\|\mathbf{x}_i\|_1 \tag{12}$$

Many l_1-minimization algorithms have been proposed in the literature [17] to solve equation (12). One can use any of those algorithms for the sparse coding, whereas we have used homotopy algorithm [17] [14].

C. Dictionary Learning

During dictionary learning step, the sparse code $\mathbf{X}$ is fixed and the cost function is optimized w.r.t to $\mathbf{D}$. Each atom of the dictionary $\mathbf{D}$ is updated in an iterative manner. The problem can be reformulated as

$$\left\langle \mathbf{d}^j \right\rangle = \underset{\mathbf{d}^j}{\arg\min}\, \rho\left(\mathbf{y}_i^{new} - \mathbf{D}^{new}\mathbf{x}_i\right)$$

The maximum likelihood solution of the above equation is given as

$$\left\langle \mathbf{d}^j \right\rangle = \underset{\mathbf{d}^j}{\arg\min} \left\| \mathbf{W}^{1/2}\left(\mathbf{y}_i^{new} - \mathbf{D}^{new}\mathbf{x}_i\right) \right\|_2^2 \tag{13}$$

where $\mathbf{d}^j$ is the j^{th} column of the dictionary $\mathbf{D}^{new}$, i.e. $\mathbf{D}^{new} = [\mathbf{d}^1, \mathbf{d}^2, ..., \mathbf{d}^k]$ and $\mathbf{y}_i^{new}$ is the probe image. To solve (13), let us rewrite it as

$$\left\langle \mathbf{d}^j \right\rangle = \underset{\mathbf{d}^j}{\arg\min} \left\| \mathbf{\Phi}_i - \bar{\mathbf{w}}_i \mathbf{d}^j \right\|_2^2 \tag{14}$$

where $j = 1, 2, .., K$ and

$$\begin{aligned} \mathbf{\Phi}_i &= \mathbf{W}^{1/2}\left(\mathbf{y}_i^{new} - \sum_{p \ne j}^{K} \mathbf{d}^p \mathbf{x}_i^p \right) \\ \bar{\mathbf{w}}_i &= \mathbf{x}_i^j \mathbf{W}^{1/2} \end{aligned} \tag{15}$$

The solution of equation (14) can be obtained by evaluating the derivative w.r.t $\mathbf{d}^j$ as given follows:

$$2\bar{\mathbf{w}}_i^T\left(\mathbf{\Phi}_i - \bar{\mathbf{w}}_i\mathbf{d}^j\right) = 0$$

$$\mathbf{d}^j = \left(\bar{\mathbf{w}}_i^T\bar{\mathbf{w}}_i\right)^{-1}\bar{\mathbf{w}}_i^T\mathbf{\Phi_i} \tag{16}$$

After the solution to equation (13) is obtained using equation (16), the dictionary atom is updated as below:

$$\mathbf{D}^{new}(:, j) = \mathbf{d}^j \tag{17}$$

The above procedure is repeated until all the dictionary atoms get updated.

D. Classification

The matrix $\mathbf{A}$ has been normalized jointly with $\mathbf{D}$ during the optimization procedure. i.e.$\forall j, \left\|\mathbf{d}^j; \mathbf{a}^j\right\|_2 = 1$. Now the matrices $\mathbf{D}, \mathbf{A}$ can be get back using the updated dictionary as

$$\begin{aligned} \tilde{\mathbf{D}} &= \left[\tilde{\mathbf{d}}^1, \tilde{\mathbf{d}}^2, ..., \tilde{\mathbf{d}}^k\right] = \left\{\frac{\mathbf{d}^1}{\|\mathbf{d}^1\|_2}, \frac{\mathbf{d}^2}{\|\mathbf{d}^2\|_2} .... \frac{\mathbf{d}^k}{\|\mathbf{d}^k\|_2}\right\} \\ \tilde{\mathbf{A}} &= \left[\tilde{\mathbf{a}}^1, \tilde{\mathbf{a}}^2, ..., \tilde{\mathbf{a}}^k\right] = \left\{\frac{\mathbf{a}^1}{\|\mathbf{d}^1\|_2}, \frac{\mathbf{a}^2}{\|\mathbf{d}^2\|_2} ...... \frac{\mathbf{a}^k}{\|\mathbf{d}^k\|_2}\right\} \end{aligned} \tag{18}$$

For classification, the multivariate ridge rgression model [10] is considered with quadratic loss and l_2-norm regularization.

$$\mathbf{W_c} = \underset{\mathbf{W_c}}{\arg\min} \|\mathbf{H} - \mathbf{W_c X}\|^2 + \lambda_1 \|\mathbf{W_c}\|_2^2$$

which yields the following solution:

$$\mathbf{W_c} = \mathbf{G X}^t \left(\mathbf{X X}^t + \lambda_1 \mathbf{I}\right)^{-1} \tag{19}$$

where $\mathbf{H}$ is the label matrix of the probe images and $\mathbf{X}$ is the sparse code matrix obtained from training data. Now for a test image $\mathbf{y}_i$, firstly compute its sparse code vector $\tilde{\mathbf{x}}_i$ and the label l of the test image $\mathbf{y}_i$ is obtained as

$$l = \underset{l}{\arg\max} \ \mathbf{W_c} \times \hat{\mathbf{x}}_i \tag{20}$$

Algorithm 1 : Robust Label Consistent Dictionary Learning

Initialization: Calculate $\mathbf{D}^{(0)}, \mathbf{A}^{(0)}$

* Obtain the initial dictionary $\mathbf{D}^{(0)} \in \mathbf{R}^{n \times K}$ by concatenating the class specific dictionaries of each class obtained using the k-svd.
* Evaluate the sparse code $\mathbf{X}^{(0)} \in \mathbf{R}^{K \times N}$ for Y and then compute $\mathbf{A}^{(0)} \in \mathbf{R}^{n \times K}$ using the below equation, which is obtained using the ridge regression model

$$\mathbf{A} = \mathbf{Q X}^t \left(\mathbf{X X}^t + \lambda_2 \mathbf{I}\right)^{-1}$$

Initialize $\mathbf{Y}_{new} = [\mathbf{Y}; \mathbf{Q}], \mathbf{D}_{new} = [\mathbf{D}; \mathbf{A}]$, $\mathbf{Y}, \mathbf{Q} \in \mathbf{R}^{n \times N}$

while not converge **do**
  Sparse coding stage: Update $\mathbf{X}$
  for i=1:N **do**
   Obtain $\mathbf{x}_i$ by solving the equation (12)
  end for
  Dictionary update Stage stage: Update $\mathbf{D}$
  for j=1:K **do**
   for i=1:N **do**
    Calculate $\mathbf{\Phi}_i, \ \bar{\mathbf{w}}_i$ using equation (15)
   end for
  Obtain $\mathbf{d}^j = \left(\bar{\mathbf{w}}_i^T \bar{\mathbf{w}}_i\right)^{-1} \bar{\mathbf{w}}_i^T \mathbf{\Phi_i}$
  update the dictionary $\mathbf{D}^{new}(:, j) = \mathbf{d}^j$
  end for
end while

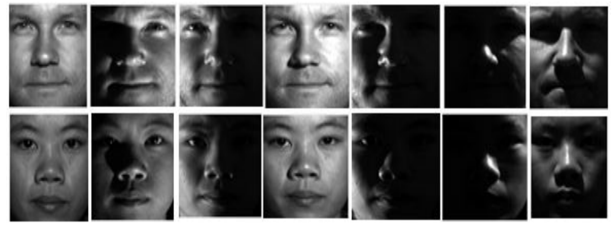

Fig. 1: Extended YaleB Database

Methodology (Training Images)	Accuracy
KSVD (15 per subject)	93.1
D-KSVD (15 per subject)	94.1
SRC (15 per subject)	80.5
LC-KSVD1 (15 per subject) [10]	94.5
LC-KSVD2 (15 per subject) [10]	95
Proposed RLC-DL (15 per subject)(logistic loss)	**95.1**
Proposed RLC-DL (15 per subject)(Huber loss)	**96.3**

TABLE I: Recognition Results for the Extended YaleB Database

IV. Experimental Results and Discussions

The proposed algorithm is evaluated on two publicly available databases namely Extended YaleB database [18] and AR database [19]. For dimensionality reduction of the face images, the standard setting in [10] [4] is followed. The random projection method [20] projects high dimensional face data onto low dimensional subspace.

A. The Extended YaleB Database

This database [18] contains 2414 grey scale images of 38 persons taken in two sessions with pose, illumination and expression variations. Each subject has nearly 64 images and each image is cropped to 192×168. The random projection method [20] has been employed for dimensionality reduction [10] of the face features. A random matrix following normal distribution with zero mean is generated and l_2-normalized. Now for random projection, the face features are projected onto 504-dimensional vector using the random matrix generated. The dictionary is learned using 15 images from each subject, so that the dictionary will have 570 dictionary atoms. For each person, a randomly selected half of the images are given for training and the other half are used for testing [10] [9]. The constants are empirically chosen as $\tau = 0.8, C_{\mu\delta} = 8$ for logistic loss and $c = 1.37$ for huber loss function respectively. Few of the images are depicted in Fig.1, and the accuracy rates are tabulated in Table I.

B. The AR Database

This database contains colour images of 126 persons taken in two sessions [19]. Each person has a total of 26 images, each of resolution (165×120) and with occlusion, disguise and variations in pose, expression, illumination from the two sessions. A subset of 2600 face images of 100 persons (50 female and 50 male) have been considered for this experiment as per the standard setting in the literature [10] [9]. The random projection

Fig. 2: AR Database

Methodology (Training Images)	Accuracy
KSVD (5 per subject)	86.5
D-KSVD (5 per subject)	88.5
SRC (5 per subject)	66.5
LC-KSVD1 (5 per subject) [10]	92.5
LC-KSVD2 (5 per subject) [10]	93.7
Proposed RLC-DL (5 per subject)(logistic loss)	**94.6**
Proposed RLC-DL (5 per subject)(Huber loss)	**95.7**

TABLE II: Recognition Results for the AR Face Database

method has been employed for dimensionality reduction of the face features. A random matrix following normal distribution with zero mean is generated and l_2-normalized. Now for random projection, the face features are projected onto 540-dimensional vector using the random matrix generated. The dictionary is learned using 5 images from each subject, so that the dictionary will have 500 dictionary atoms. For each person, a randomly selected 20 images are given for training and the remaining 6 images are used for testing [10] [9]. The constants are empirically chosen as $\tau = 0.8, C_{\mu\delta} = 8$ for logistic loss and $c = 2$ for huber loss function respectively. Few of the images from this data set are depicted in Fig.2, and the accuracy rates are tabulated in Table II

V. Conclusion

An efficient dictionary learning approach namely Robust Label Consistent Dictionary Learning (RLC-DL) algorithm is proposed and evaluated for sparse coding and dictionary learning in a robust formulation. Introduction of the discriminative sparse code error into the cost function and assigning the weights in an adaptive and iterative way to the pixels based on their fiedility error made the RLC-DL algorithm to robustly differentiate between the outliers and inliers. The weight functions from the theory of robust estimators are chosen in such a way that, effect of outliers gets diminished in the coding process. The presence of discriminative sparse code error forces the sparse codes from similar class to have similar sparse codes and sparse codes from different classes to be discriminant. A simple linear classifier is used for classification. The proposed RLC-DL method yielded promising results on two publicly available databases with considerable variations in pose, illumination, expressions and occlusion/disguise. The experimental results clearly demonstrated that RLC-DL outperforms the algorithm SRC [4], KSVD [7], D-KSVD [9] and LC-KSVD [10].

References

[1] R. Rubinstein, A. M. Bruckstein, and M. Elad, "Dictionaries for sparse representation modeling," *Proceedings of the IEEE*, vol. 98, no. 6, pp. 1045–1057, 2010.

[2] M. Elad and M. Aharon, "Image denoising via sparse and redundant representations over learned dictionaries," *IEEE Transactions on Image processing*, vol. 15, no. 12, pp. 3736–3745, 2006.

[3] Y. Xu, Z. Li, J. Yang, and D. Zhang, "A survey of dictionary learning algorithms for face recognition," *IEEE Access*, 2017.

[4] J. Wright, A. Y. Yang, A. Ganesh, S. S. Sastry, and Y. Ma, "Robust face recognition via sparse representation," *IEEE transactions on pattern analysis and machine intelligence*, vol. 31, no. 2, pp. 210–227, 2009.

[5] M. Elad, "Sparse and redundant representation," 2010.

[6] M. Yang, L. Zhang, X. Feng, and D. Zhang, "Fisher discrimination dictionary learning for sparse representation," in *Computer Vision (ICCV), 2011 IEEE International Conference on*. IEEE, 2011, pp. 543–550.

[7] M. Aharon, M. Elad, and A. Bruckstein, "k-svd: An algorithm for designing overcomplete dictionaries for sparse representation," *IEEE Transactions on signal processing*, vol. 54, no. 11, pp. 4311–4322, 2006.

[8] R. Rubinstein, M. Zibulevsky, and M. Elad, "Efficient implementation of the k-svd algorithm using batch orthogonal matching pursuit," *Cs Technion*, vol. 40, no. 8, pp. 1–15, 2008.

[9] Q. Zhang and B. Li, "Discriminative k-svd for dictionary learning in face recognition," in *Computer Vision and Pattern Recognition (CVPR), 2010 IEEE Conference on*. IEEE, 2010, pp. 2691–2698.

[10] Z. Jiang, Z. Lin, and L. S. Davis, "Label consistent k-svd: Learning a discriminative dictionary for recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 35, no. 11, pp. 2651–2664, 2013.

[11] M. Yang, L. Zhang, X. Feng, and D. Zhang, "Sparse representation based fisher discrimination dictionary learning for image classification," *International Journal of Computer Vision*, vol. 109, no. 3, pp. 209–232, 2014.

[12] M. Yang, L. Zhang, J. Yang, and D. Zhang, "Robust sparse coding for face recognition," in *Computer Vision and Pattern Recognition (CVPR), 2011 IEEE Conference on*. IEEE, 2011, pp. 625–632.

[13] W. Deng, J. Hu, and J. Guo, "Extended src: Undersampled face recognition via intraclass variant dictionary," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 34, no. 9, pp. 1864–1870, 2012.

[14] C.-P. Wei and Y.-C. F. Wang, "Undersampled face recognition via robust auxiliary dictionary learning," *IEEE Transactions on Image Processing*, vol. 24, no. 6, pp. 1722–1734, 2015.

[15] S. Boyd and L. Vandenberghe, *Convex optimization*. Cambridge university press, 2004.

[16] R. Maronna, R. D. Martin, and V. Yohai, *Robust statistics*. John Wiley & Sons, Chichester. ISBN, 2006.

[17] A. Y. Yang, S. S. Sastry, A. Ganesh, and Y. Ma, "Fast 1-minimization algorithms and an application in robust face recognition: A review," in *Image Processing (ICIP), 2010 17th IEEE International Conference on*. IEEE, 2010, pp. 1849–1852.

[18] A. S. Georghiades, P. N. Belhumeur, and D. J. Kriegman, "From few to many: Illumination cone models for face recognition under variable lighting and pose," *IEEE transactions on pattern analysis and machine intelligence*, vol. 23, no. 6, pp. 643–660, 2001.

[19] A. M. Martinez, "The ar face database," *CVC technical report*, vol. 24, 1998.

[20] N. Goal, G. Bebis, and A. Nefian, "Face recognition experiments with random projection," in *Proceedings SPIE Vol*, vol. 5779, 2005, pp. 426–437.