Low Rank Group Sparse Representation based Adaptive Face Recognition

Shivangi Yadav

IIIT-D-MTech-CS-GEN-14-104

May 5, 2016

Indraprastha Institute of Information Technology, Delhi
New Delhi

<u>Thesis Advisors</u>

Dr. Mayank Vatsa
Dr. Richa Singh

Submitted in partial fulfillment of the requirements
for the Degree of M.Tech. in Computer Science

Keywords : Domain Adaptation, Dictionary, Group Sparse, Low Rank

Certificate

This is to certify that the thesis titled **"Low Rank Group Sparse Representation based Adaptive Face Recognition"** submitted by **Shivangi Yadav** for the partial fulfillment of the requirements for the degree of *Master of Technology* in *Computer Science & Engineering* is a record of the bonafide work carried out by her under our guidance and supervision at Indraprastha Institute of Information Technology, Delhi. This work has not been submitted anywhere else for the reward of any other degree.

Dr. Mayank Vatsa

Dr. Richa Singh

Indraprastha Institute of Information Technology, Delhi

Abstract

Face recognition is an important area of research due to its requirement in our day-to-day life, be it surveillance or authentication. Current advancements in technology and computational power have shown promising results to solve this problem. Despite such furtherance, face recognition under uncontrolled environment still remains a challenging task and many state-of-the-art algorithms are unable to serve the purpose due to several challenges including varying illumination, pose, resolution, and occlusion. One primary reason for low performance is difference in training and testing data distribution. In this research, we propose an algorithm for face identification with varying pose and illumination. We propose an adaptive dictionary learning framework with Group Sparse Representation based Classifier to learn domain invariant dictionary representation of the given data. The algorithm adapts the representation learnt from the source domain with respect to the target domain in order to reduce the differences arising due to changes in the training and testing data distributions. Further, the data may contain noise and affect the dictionary atoms and group sparse coefficients, thereby hindering the discriminative power of the learnt dictionary. We propose to solve this problem using low rank minimization on dictionary atoms and group sparse coefficients. The effectiveness of the proposed algorithm is evaluated on the CMU MultiPIE and Extended YaleB face datasets for varying pose and illumination.

Acknowledgments

I wish to express my sincere gratitude towards all those people who contributed towards my Masters Thesis in many ways. Firstly, I would like to thank my advisors Dr. Mayank Vatsa and Dr. Richa Singh. Their humble support and guidance gave me confidence and helped me prosper in many ways. They supported me and my work at all stages and was their to guide me throughout the journey. They pushed me everytime so that I can work to my full potential and gain knowledge that not only helped me towards my thesis completion but also motivated me towards various research areas under this domain. It's been my pleasure to work under their guidance. I would also like to mention Dr. Angshul Majumdar and Maneet Singh, without whose support this work would not have been what it is today. This section is not complete without a thank you note to academic department for their help and never ending support.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Overview and Research Motivation

Face has been established as one of the least invasive biometric modalities, thereby, making it one of the most well explored signatures for person identification. It provides discriminative textural and structural information, which is often used for identity recognition. Many algorithms have been proposed to automate this task under several covariates such as varying resolution, occlusion and disguise [12, 39]. Though recent algorithms claim high accuracies [25], the performance of the same in real-world conditions is still an open research problem.

One of the major challenges associated with automated face recognition in completely unconstrained scenarios is the presence of pose and illumination variations. Algorithms that utilize only frontal, well-illuminated face images for learning a classification model are often ineffective in the presence of such covariates. This is primarily because the data distribution on which the classifier is trained might differ from the distribution of the test samples. Studies have shown that any variation in illumination or viewing angle can affect the performance of a recognition system significantly. For example, Figure 1.1 and 1.2 depicts images from a single individual with varying pose and illumination. Illumination variations such as capturing photos in indoor or outdoor can cause traditional face recognition algorithms to fail.

Multiple features can be used in face recognition algorithms such as geometric, appearance,

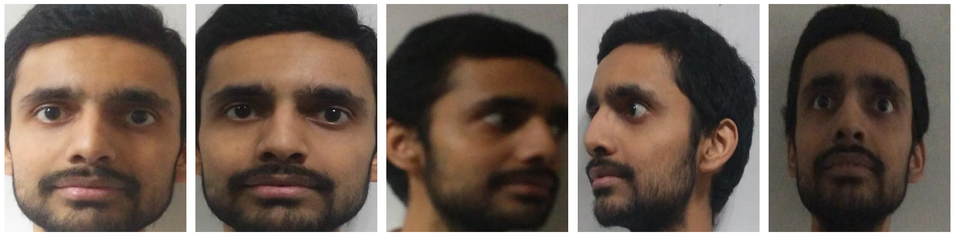

Figure 1.1: Images captured under unconstrained environment for Subject-1

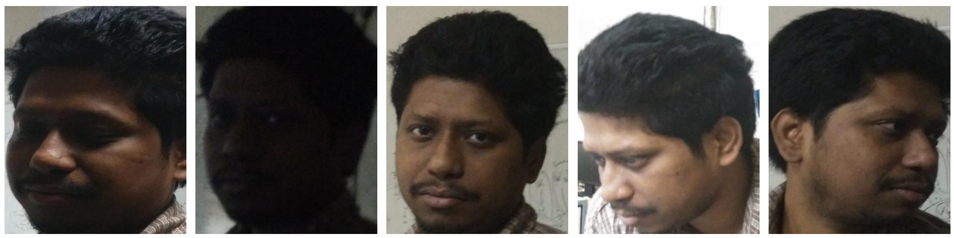

Figure 1.2: Images captured under unconstrained environment for Subject-2

texture and representative learning based approaches. Geometry based approaches tend to use relative position of facial features such as eyes, nose, chin to get appropriate features while appearance based features rely on low dimension representation of image itself e.g. Eigenfaces and SLAM. Although geometry based features are more resilient towards small variation of pose and illumination variation, they are not very effective in unconstrained facial recognition. Appearance based methods perform on these condition only when they are provided with large amount of training samples for all possible variations, which itself is a challenging task. Texture based approaches use handcrafted features Local Binary Features (LBP) and Scale Invariant Feature Transform (SIFT) whereas, representative learning approaches are data driven that include deep learning and dictionary learning. Recently, 3-D modeling of face images constructed with the training data have generated a lot of interest [20]. These 3-D models can be used to render various illumination and pose variations while evaluating any test sample; however, 3-D modeling of face also requires large number of samples.

1.2 Literature Review

The major approaches for face recognition under pose and illumination variation include methods like 3-D face reconstruction, image mosaicing, deep learning and domain adaptation [2, 9, 15, 24, 30, 32]. In this section, we present a brief overview on some existing approaches to deal with the above mentioned problem:

- Passils *et al.* [20] proposed a novel 3-D face recognition method that takes into account facial symmetry to handle all pose variations in the data. It uses an automatic landmark detector that aims to detect occluded areas and pose for each given face image. Using facial symmetry, an annotated face model is learnt and fitted to the facial scan to overcome the challenges of missing data. The resultant image is pose invariant and thus helps in overcoming the challenges of pose variations. Passils proposed algorithm, unlike other face recognition methods, aims to perform comparisons between interpose scans with the help of wavelet based biometric signature. The proposed method was evaluated on databases from the University of Houston and the University of Notre Dame. It obtained an average rank-one recognition rate of 83.7 percent.

- Zhu *et al.* [38] proposed to tackle the problem of pose and illumination variations by learning Face Identity Preserving features (FIP). The Authors designed a deep network consisting of a feature extraction and a reconstruction layer. The former encodes an image into FIP features and the extracted features are then used to reconstruct them into a canonical view by latter. Also, unlike conventional descriptors, FIP features aim to reduce intra-class variances while still maintaining discriminative properties between different classes. These properties of FIP features makes it possible to improve performance of descriptors like LBP and Gabor on pose and illumination variations. LBP and Gabor descriptors are learned from reconstructed face image in canonical view, thereby, eliminating the variations. The proposed algorithm was tested on CMU MultiPIE face database.

- Singh *et al.* [29] proposed to overcome the problems of pose variation using face mosaicing algorithm to generate a composite face image using frontal and semi-profile images. This is achieved using a terrain transform that determines the transformation relating the side and frontal profiles by exploiting the neighborhood properties. Further, multi-resolution splining is used to blend the profiles to generate a composite face image, template face mosaic. A given input image is then matched with the template face mosaics present in the gallery using Local Binary Pattern matching algorithm. The proposed algorithm was evaluated on a dataset of 27 users and results showed the effictiveness of the algorithm.

- Sparse representation and dictionary learning algorithms has shown some interesting results for face recognition. Using this, Yang *et al.* [33] proposed a dictionary learning method (DL) to achieve improved pattern classification performance. A class-wise structured dictionary is learned using Fisher discrimination criteria such that the reconstruction error can be used for pattern classification. Fisher discrimination criteria is being imposed on the coding coefficients in order to maximize between-class scatter but at the same time minimize the within-class scatter. This is then known as Fisher discrimination DL (FDDL) method that uses both sparse coding coefficients and the discriminative information in the reconstruction error. The proposed algorithm is then evaluated on MultiPIE and Yale face datasets.

- Sharma *et al.* [27] discussed in their paper that an algorithm dealing with domain adaptation should have following properties: supervised, generalizable to unseen classes, ability to handle multi-view and not domain dependent. Hence, he proposed Generalized Multi-view Analysis (GMA) that aims to learn a common discriminative subspace and abide all the above mentioned properties. It solves a joint, relaxed Quadratic Constrained Quadratic Problem (QCQP) over different feature spaces and obtains a single subspace i.e., it solves a generalized eigenvalue problem leading to a globally optimal solution. The paper discusses that GMA is a generalized extension of Canonical Correlation Analysis, Partial Least Squares and Bilinear Model. GMA is tested for simultaneous pose and illumination variation on CMU MultiPIE face dataset and Wiki text-image data.

- Shekhar *et al.* [28] discussed that data driven dictionaries perform very well on various classification tasks. Thus, they proposed to represent the source as well as target domain data by a common dictionary such that the difference between two domains is minimized. Since the features from the two domains may not be well connected, therefore, they projected data from both domains, separately, on a low dimensional space. This helps in handling any changes in feature type and dimensions across the domains. Simultaneously, a common dictionary is learnt that represents the projected data from both domains. This joint optimization offers generalizability, efficiency and helps to find common internal structure between source and target domain that is well represented by linear sparse combinations of dictionary atoms. This algorithm has been evaluated on CMU MultiPIE face dataset, as well as, Caltech-256 and Amazon datasets for object recognition.

- Qui *et al.* [23] discussed a domain adaptation method to learn to incrementally adapt a dictionary from source to target domain such that the difference between the two domains can be minimized. They proposed a Domain Adaptive Dictionary Learning framework (DADL) that learns a transformation for dictionary in one domain to another domain, while taking care of the domain invariant sparse codes of a signal. Dictionary atoms and domain invariant sparse codes are jointly learnt by solving an optimization problem. This algorithm was evaluated on CMU PIE and extended YaleB dataset.

- Sharma *et al.* [26] introduced a view-based subspace, hybrid-eigenfaces, to generate images for different illumination and poses from a single image. Hybrid-eigenfaces are very different from view-based subspace as proposed by Pentland and Turk [21]. For example Hybrid-eigenfaces provides high co-relation for a subject even under different illumination and poses, while this advantage is not offered by view-based eigenfaces. Hybrid eigenfaces are then further combined with Global Linear Regression, proposed by Chai *et al.* [3], to generate 2D images with different pose and illumination. These images were then used for training for some state-of-art methods on FERET and extended YaleD face dataset to compare its robustness and generalization.

Table 1.1: Brief survey on Literature Review

Paper	2D or 3D	Dataset	Technique	Average Accuracy
Passils *et al.* [20]	3D	3-D face dataset from University of Houston and University of Notre Dame	Automatic landmark detection and Wavelet transformation	83.7
Zhu *et al.* [38]	2D	CMU MultiPIE	Deep Network	95.6
Singh *et al.* [29]	2D	WVU Visible Face dataset, WVU SWIR Face dataset and CMU MultiPIE	Face mosaic	97.5, 98.2 and 96.9
Yang *et al.* [33]	2D	CMU MultiPIE and extended YaleB	Class-wise structured dictionary using Fisher discrimination criteria	93.9 and 91.9
Sharma *et al.* [27]	2D	CMU MultiPIE and Wiki text image data	Generalized Multiview Analysis	99.2 and 99.7
Shekhar *et al.* [28]	2D	CMU MultiPIE, Caltech-256 and Amazon	Dictionary Learning on a low dimensional space	98.5 and 46.2
Qui *et al.* [23]	2D	CMU MultiPIE and extended YaleB	Domain Adaptive dictionary learning framework	90.4
Sharma *et al.* [26]	2D	FERET and extended YaleB	Hybrid eigenface with Global Linear Regression	76 and 61.32

1.3 Research Contributions

In this research, a Low Rank Group Sparse Representation based Classifier(LR-GSRC) is proposed for face recognition with pose and illumination variations. The algorithm is built upon existing Group Sparse Classifier [10] and utilizes incremental learning [23] with trace norm regularizer [17] for addressing the given problem. In Dictionary Learning approaches, images are represented as a linear combination of atoms of a dictionary. Generally, for a given dictionary, the total number of atoms are large as opposed to the atoms used for the reconstruction of a given image, which results in sparse coefficients for an image. Recently, Group Sparse Classifier has been proposed which assumes that a test sample can be represented as a linear combination of training samples belonging to the same group as that of the given test sample. Since the samples are linearly correlated, the dictionary for a particular group should fall in a low dimensional manifold [14]. To enforce this, a trace norm regularizer on the group-wise dictionaries is introduced in the dictionary learning algorithml. As mentioned earlier, since the distribution of the test samples (target domain) might differ from the distribution of the training samples (source domain), the above framework is learnt in an incremental manner. The research work is evaluated on CMU MultiPIE [11] and Extended YaleB [8] face dataset for varying pose and illumination.

Chapter 2

Preliminaries

This chapter provides some preliminaries that is required for the proposed algorithm LR-GSRC. Details about Sparse Approximation, Dictionary Learning, Group Sparse Representation based Classifier (GSRC) and Low Rank Minimization are given below.

2.1 Sparse Approximation

Sparse approximation is a technique used to obtain a vector, sparse in nature, which is an approximate solution for a system of equations. It has found wide use in the applications of image processing, mainly because of the known fact that signals and images can be sparse in some dictionary D [7, 13, 31, 34].

Given a linear system of equations, $y = \mathrm{D}x$, where D is an undetermined matrix also known as dictionary and x is a signal that has to be estimated with the constraint that it should be sparse in nature. The motivation towards finding a sparse representation of an input signal is that even though the observed values of input y are in high dimension space but the actual signal has been organized in a low dimensional subspace [4, 5]. Sparsity in simple terms implies that very few components of x are non-zero. Therefore, the input signal y can be decomposed into a linear combinations of few vectors in D, known as atoms. Mathematically, the sparse approximation problem is represented as,

$$\underset{x}{\operatorname{argmin}} \|x\|_o \, subject \ to \ y = Dx \tag{2.1}$$

where $\|x\|_o$ refers to l_o norm that gives the number of nonzero entries in vector x. This is a NP-Hard problem that can be reduced to NP-complete subset selection problems. Often the observed values of input signal y are noisy in nature. For such cases, the sparse approximation problem can be represented as:

$$\underset{X}{\operatorname{argmin}} \|Y - DX\|_2^2 + \lambda \|\alpha\|_1 \tag{2.2}$$

where, $||\alpha||_1$ induces sparsity and λ is a slack variable that is responsible for balancing the trade-off between finding a sparse representation and fitting the data properly.

Several algorithms have been proposed to solve sparse approximation problem. The most commonly used techniques are:

- Matching Pursuit: It is a greedy algorithm using iterative approach to solve l_o problem mentioned in equation 2.1. Matching pursuit aims to find a basis vector D such that it maximizes the correlation with the residual and then re-computes the coefficients and residual by projecting them on all atoms of the dictionary.

- Orthogonal Matching Pursuit: Matching Pursuit has a drawback that an atom can be picked multiple times that is overcome by Orthogonal Matching Pursuit algorithm. It is very much similar to Matching Pursuit and in addition maintains that an atom picked once will not be picked again. It maintains an active set of atoms that have already been picked and adds a new atom after every new iteration. Similar to Matching Pursuit the residual is projected onto this linear combination of active atoms such that an orthogonal recomputed residual is obtained.

- Projected Gradient Descent: This method works in a similar manner as the Gradient

Descent i.e., given an initial point it provides the information that point to new search directions. Since aim is to sparse solution therefore, the presumptive solutions are sparsely projected onto scaffold of k-vectors.

- LASSO: This algorithm aims to solve the l_1 minimization problem given in equation 2.1. Unlike Matching Pursuit, instead of projecting the residual on dictionary atoms, it aims to move the residual iteratively towards the atoms by a small step each time.

2.2 Dictionary Learning

In dictionary learning algorithms, images are represented as a linear combination of atoms of a dictionary. Given a signal y and dictionary D, sparse representation of y can be learned through following optimization problem:

$$\hat{x} = \underset{x}{\operatorname{argmin}} \|x\|_o \, subject \ to \ y = Dx \tag{2.3}$$

where, $\|x\|_o$ refer to l_o norm that gives the number of nonzero entries in vector x. Recently, many new approaches have been discussed to learn an efficient dictionary [15, 16] from the given data. It has mainly been influenced by recent advances in sparse algorithms and representation theory. One of the established methods of learning a dictionary from training samples is the K-SVD algorithm [1, 36]. K-SVD is a dictionary learning algorithm, which is a generalization of k-means clustering method. K-means clustering can also be regarded as a method for sparse representation that aims to find the best possible codes to represent the input signal y by nearest neighbor method through solving following equation:

$$\underset{D,X}{\operatorname{argmin}} \|Y - DX\|_F^2 \quad s.t. \vee_i, \ \|x_i\|_o = 0 \tag{2.4}$$

The sparse representation $\|x\|_o = 1$ enforces the K-means algorithm to have just one atom in the dictionary. Since K-SVD algorithm aims to achieve linear combinations of atoms, therefore, the constraint is updated in such a way that x_i is greater than 1, but smaller than a threshold T_o,

$$\underset{D,X}{\operatorname{argmin}} \|Y - DX\|_F^2 \quad s.t.\ \vee_i,\ \|x_i\|_o \leq T \tag{2.5}$$

where X = $[x_1,...,x_N]$, $x_i \ \epsilon \in R^k$ are sparse codes of N input signals Y, D = $[d_1,...,d_k]$, $d_i \ \epsilon \in R^n$ and T restricts the signal to have less than T items in its decomposition. k represents the number of atoms in learnt dictionary and n represents the number of samples on which the dictionary has been learned. This equation is solved using Orthogonal Matching Pursuit (OMP) to learn both, sparse codes and dictionary alternatively as well as iteratively. At the end, result is a sparse representation that very well represents the input signal y.

2.3 Group Sparse Representation based Classification

2.3.1 Sparse Representation Based Classification

Sparse Representation based Classification (SRC) [31] assumes that a test sample can be represented as a linear combination of training samples belonging to the same class as the test sample. For example: if v_{test} is a sample belonging to the k^{th} class then, it can be depicted as,

$$v_{test} = \alpha_{k,1}v_{k,1} + \alpha_{k,2}v_{k,2} + ... + \alpha_{k,n}v_{k,n} + \epsilon \tag{2.6}$$

Where, v_{test} belongs to class k, $v_{i,k}$ represents i^{th} training sample from k^{th} class and ϵ is the approximation error. Since the correct class of v_{test} is not known, therefore, for classification purpose SRC represents v_{test} as linear combination of all samples from all classes:

$$v_{test} = V\alpha + \epsilon \tag{2.7}$$

where, V = [$[V_{1,1}.....V_{1,N}]$, $[V_{2,1}.....V_{2,N}]$, , $[V_{C,1}.....V_{C,N}$]] and α = [$[\alpha_{1,1}.....\alpha_{1,N}]$, $[\alpha_{2,1}.....\alpha_{2,N}]$, , $[\alpha_{C,1}.....\alpha_{C,N}$]]. Here $V_{i,j}$ represents a j^{th} training sample from class i, C depicts the number of classes in the given data and N represents the number of training samples belonging to a class. As mentioned earlier, a given test sample can be linearly associated with only the training samples belonging to the same class as the test sample and samples from other classes should not contribute. Therefore, α is likely to be sparse in nature i.e., it will have a zero value for all other classes and non-zero values only for the class to which it belongs. This can be solved using following minimization problem:

$$\min_{\alpha} \|v_{test} - V\alpha\|_2^2 + \lambda \|\alpha\|_1 \tag{2.8}$$

Majumdar *et al.* [19] and Elhamifar *et al.* [6] claim that l_1-norm does not explicitly impose that α for correct class should be non-zero and otherwise zero. Instead, it can better be enforced using supervised $l_{2,1}$-norm. So the minimization changes to:

$$\min_{\alpha} \|v_{test} - V\alpha\|_2^2 + \lambda \|\alpha\|_{2,1} \tag{2.9}$$

This is known as block/joint SRC and works well for simple classification problems [19, 35], however, yields very low performance for face recognition in comparison to SRC. Using equation 2.9 Wright *et al.* [31] proposed the following algorithm in order to determine class of a given test sample:

- For each class c, reconstruct a sample $v_{recon}(c)$ by the linear combination of training samples from that class:

$$v_{recon}(k) = V_k \alpha_k \tag{2.10}$$

- Find the error between a given test sample and the reconstructed sample

- Assign the test sample to the class with the minimum reconstruction error

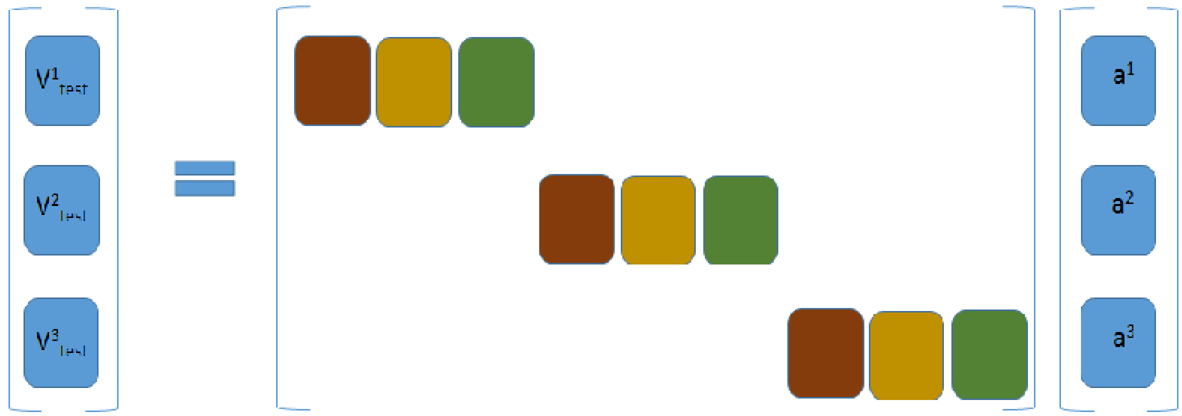

Figure 2.1: Group Sparse Representation based Classifier

2.3.2 Group Sparse Representation based Classification

Group Sparse Representation based Classification [10] is similar to SRC, however, it aims to handle multiple features at same time i.e., if there are N modalities then, for each modality SRC model holds true (test sample from a modality can be linearly associated with only the training samples belonging to the same modality and class as test sample),

$$v^i_{test} = V^i \alpha^i + \epsilon \tag{2.11}$$

$$v^i_{test} = \alpha^i_{k,1} v^i_{k,1} + \alpha^i_{k,2} v^i_{k,2} + ... + \alpha^i_{k,n} v^i_{k,n} + \epsilon \tag{2.12}$$

where v^i_{test} depicts that this test sample refers to i^{th} modality, $v^i_{k,n}$ refers to n^{th} training sample from k^{th} class and i^{th} modality, and similarly, $\alpha^i_{k,n}$ refers to α coefficient for n^{th} training sample from k^{th} class and i^{th} modality. Since SRC algorithm is true for each individual modality, therefore, α^i_k is sparse in group i that contains non-zero values for samples from the i^{th} group and k^{th} class, and zero otherwise. Therefore, equation 8 can now be solved using:

$$\min_{Z} \|v_{test} - V\alpha\|^2_2 + \lambda \|\alpha\|_{2,1} \tag{2.13}$$

The classification algorithm for Group Sparse is also very similar to SRC. For a given test sample, following steps are followed to perform classification:

- For each class c and modality i, reconstruct a sample $v^{i}_{recon}(c)$ by the linear combination of training samples from that class and modality

- Find error between given test sample and reconstructed sample

- Assign the test sample to the class with the minimum reconstruction error

2.4 Low Rank Approximation

As discussed, in SRC samples belonging to the same class are linearly correlated. Therefore, an input signal can be represented through few atoms of an over-complete dictionary. Such an architecture has achieved quite impressive performance for applications like image classification. Quality of a learnt dictionary is crucial factor for sparse representation. The SRC algorithm takes entire training set to learn a dictionary, however, given a large dictionary, sparse coding is computationally expensive. Therefore, some researches [18, 22] focused on learning discriminative and compact dictionaries. The performance of algorithms on applications like image classification improves dramatically for a well-constructed dictionary. Also, compact dictionary has been proven to be efficient for encoding step. However, performance of such methods degrade if the training data is noisy or contaminated (for example lighting variations, occlusion, pixel corruption, disguise, etc.). One approach towards learning a compact and noise-free dictionary for better image classification is low rank approximation [14, 37].

Low Rank Approximation is a minimization problem that aims to find the best fit between a given matrix and an approximating matrix, subject to the constraint that it has low rank. It has wide application in the field of image compression and mathematical modeling.

2.4.1 Basic Low Rank Minimization Problem

The fit measured by Frobenius norm,

$$\underset{D}{\operatorname{argmin}} \left\| D - \hat{D} \right\|_F^2 \quad s.t.\ rank(\hat{D}) \leq r \tag{2.14}$$

contains an analytic in reference to singular value decomposition of given matrix D i.e., Let singular value decomposition of D be

$$D = U\Sigma V^T \epsilon\ \mathbb{R}^{m*n} \tag{2.15}$$

where, $U = [U_1\ U_2]$, $V = [[V_1\ V_2]$ and $\epsilon = \text{diag}(\sigma_1,..,\sigma_m)$. The rank-r matrix obtained from this singular value decomposition is:

$$\hat{D}^* = U_1 \Sigma V_1^T \epsilon\ \mathbb{R}^{m*n} \tag{2.16}$$

Such that,

$$\left\| D - \hat{D}^* \right\|_F = min_{rank(\hat{D}) \leq r} \left\| D - \hat{D} \right\|_F = \sqrt{\sigma_1^2 + ... + \sigma_n^2} \tag{2.17}$$

2.4.2 Trace Norm

The Schatten p-norm acts on the singular values of a given matrix in the following way,

$$\|D\|_p = \left(\sum_{i=1}^{min(m,n)} \sigma_i^p \right)^{1/p} \tag{2.18}$$

To obtain a low rank matrix for given matrix D. If p value in equation 2.18 is substituted with value p = 2, it yields the Frobenius norm and if the p is substituted with value p = 1, it yields the trace norm (also known as nuclear norm). So, for trace norm equation 2.18 will look like,

$$\|D\|_* = trace\left(\sqrt{D^* D}\right) = \sum_{i=1}^{min(m,n)} \sigma_i \tag{2.19}$$

As discussed before, samples belonging to same class are linearly correlated therefore, the dictionary for a particular group should fall in a low dimensional manifold. To enforce this, a trace norm regularizer on the group-wise dictionaries can be applied in order to obtain a low rank group-wise dictionary [16, 37]. One advantage of low rank representation is that since it acts directly on singular values therefore, it take cares of any noise present in the training samples.

Chapter 3

Low Rank Group Sparse Representation based Classifier (LR-GSRC)

The existing algorithms for domain adaptation on pose and illumination variations either require large amount of data to identify different poses belonging to the same identity, can handle only a specific pose in the target domain at a time or are unable to find correlation between images having left and right view of the same individual. Some of these disadvantages are overcome by proposing a novel framework based on adaptive dictionary learning and Group Sparse Representation based Classifier (GSRC). Adaptive dictionary learning approach helps interpolating the path between the source and target domain, thereby obtaining a good representation of data present in the target domain. On the other hand, GSRC trains the classifier to learn both left and right view simultaneously such that, even if the target domain contains different views for a class then also it should be able to assign it to the right identity.

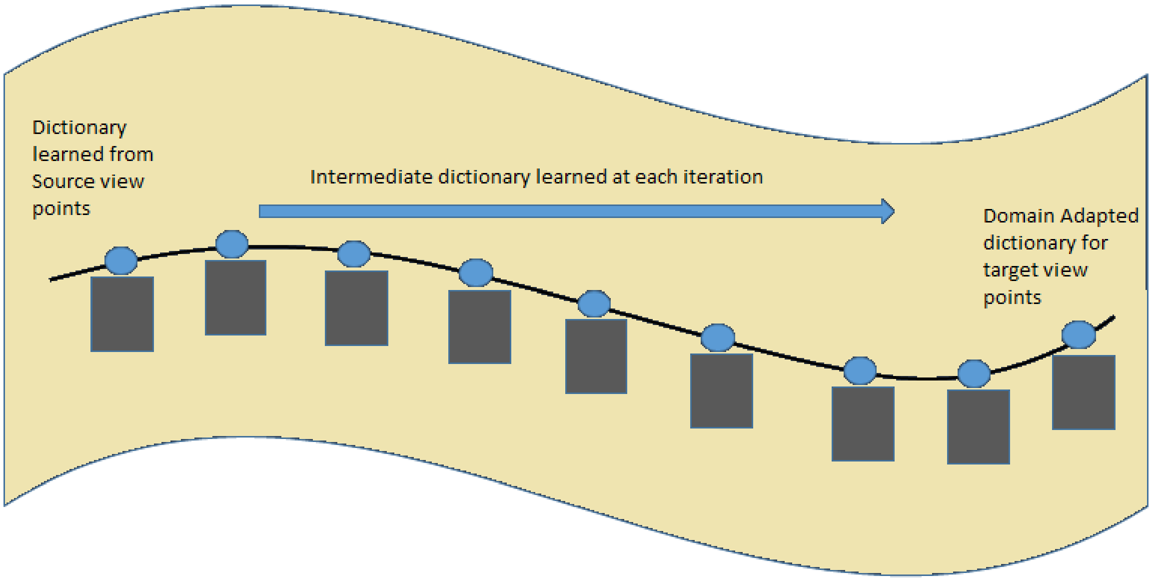

Figure 3.1: Given a source domain, proposed algorithm aims to interpolate the path between source and target domain by learning intermediate dictionaries until it finds the best possible representation for target domain

3.1 Proposed Algorithm

In this section the proposed algorithm is described in detail. Let $Y_s = [Y_{1,s}, Y_{2,s}, ..., Y_{c,s}]$ contains all samples from source domain, with a total of N_s instances from c different classes. Hence, $Y_{i,s} \: \epsilon \in R^{n*ms_i}$, where n is the dimension of the samples and ms_i refers to the i^{th} class size in the Source Domain. Similarly, $Y_t = [Y_{1,t}, Y_{2,t}, ..., Y_{c,t}]$ contains samples from target domain such that $Y_{i,t} \: \epsilon \in R^{n*mt_i}$. From Y_s, a dictionary D_j is learnt for each class. D = $[D_1, D_2, .., D_c]$, where D_j is the dictionary for j^{th} class and $D_j \: \epsilon \in R^{n*p}$. Here p depicts number of atoms in the dictionary.

For Group Sparse Classifier, let there be i groups in the source domain on which the Group Sparse coefficients are trained such that Y_t^i represents instances from target data belonging to i^{th} group and D_j^i represents dictionary from i^{th} group and j^{th} class. The aim here is to incrementally learn Group Sparse coefficients and dictionary such that at k^{th} iteration dictionary $D_{*,k}$ is closer to the target domain as compared to the $k-1^{th}$ iteration dictionary. Here, $D_{*,k}$ refers to the dictionary D = $[D_1, D_2, .., D_c]$ learned at the k^{th} iteration for all classes c in the data.

3.1.1 Training

Given Y_t and Y_s (instances from target and source domain respectively), the algorithm is as follows:

Step 1: Learn the source dictionary $D_{*,o}$ using samples from Y_s. Using this dictionary as initial point, our aim is to incrementally learn Group Sparse coefficients α and target dictionary that gives the best representation for the target domain.

Step 2: $D_{*,k}$ is updated for the next intermediate domain $k+1$ to incrementally adapt to the target data [23]. $D_{*,k+1}$ is learnt on the basis of its coherence with the dictionary in k^{th} domain and residual of instances in Y_t. The residual,$Z_{*,k}$ is obtained using the following:

$$X_{*,k} = \underset{X}{\operatorname{argmin}} \left\| Y_t - D_{*,k} X \right\|_F^2, s.t. \vee_i, \left\| p_i \right\|_o \leq T \tag{3.1}$$

$$Z_{*,k} = \left\| Y_t - D_{*,k} X_{*,k} \right\|_F^2 \tag{3.2}$$

here $X_{*,k} = [p_1, ..., p_{N_t}]$ refers to the sparse coefficients of data instances in Y_t decomposed with dictionary from k^{th} iteration. p_i refers to sparse coefficients of data instances belonging to class i. The updation in $D_{*,k}$ atoms, Δ $D_{*,k}$, to obtain $D_{*,k+1}$ is formulated using following minimization:

$$\min_{\Delta D_{*,k}} \left\| Z_{*,k} - \Delta D_{*,k} X_{*,k} \right\|_F^2 + \lambda \left\| \Delta D_{*,k} \right\|_F^2 \tag{3.3}$$

The first term is responsible for adjustments in atoms of dictionary $D_{*,k}$ in order to decrease residual reconstruction error $Z_{*,k}$. The second term is used to control sudden changes in dictionary atoms between current domain and next domain. Hence, $D_{*,k+1}$ can be formulated using:

$$\Delta D_{*,k} = Z_{*,k} X_{*,k}^T (\lambda I + X_{*,k} X_{*,k}^T)^{-1} \tag{3.4}$$

$$D_{*,k+1} = D_{*,k} + \Delta D_{*,k} \tag{3.5}$$

This step is repeated to learn intermediate representation till the best representative dictionary of target data is obtained. This is enforced by a stopping criteria: $\|\Delta D_{*,k}\|_F < \delta$.

Step 2: Using adapted dictionary with respect to Y_t, α for GSRC is learned using the following formulation:

$$\min_{D,\alpha} \left\|Y_t^i - D_{*,k}^i \alpha^i\right\|_2^2 + \lambda \left\|\alpha^i\right\|_{2,1} + \sum_j \left\|\alpha_{j,k}^i\right\|_* + \sum_j \left\|D_{j,k}^i\right\|_* \tag{3.6}$$

here, $\|a\|_*$ refers to trace norm that is used as low rank regularization on dictionary. $D_{j,k}^i$ represents dictionary for i^{th} group and j^{th} class at k^{th} iteration. This approach has been summarized in Algorithm 1.

Data: Source Dictionary $D_{*,o}$ learnt after Step 1, target data Y_t, sparsity level T,
Result: $D_{*,k}$ and α for intermediate domains
initialization;
do
  1. Obtain Obtain $Z_{*,k}$ from Y_t and $D_{*,k}$ using equation (3.1) and (3.2)
  2. Update atoms in $D_{*,k}$ to get next intermediate domain $D_{*,k+1}$ using (3.3), (3.4) and (3.5);
while $\|\Delta D_{*,k}\|_F < \delta$;
3. Learn Group Sparse coefficients α with adapted Dictionary using equation (3.6)

Algorithm 1: Low Rank GSRC

3.1.2 Testing

For a given test sample, following steps are followed:

1. For each class c, reconstruct a sample $v_{recon}(c)$ by the linear combination of training samples from that class:

$$v_{recon}(k) = V_k \alpha_k \tag{3.7}$$

2. Find error between given test sample and reconstructed sample

3. Assign the test sample to the class with the minimum reconstruction error

Chapter 4

Experiments and Results

The performance of the proposed algorithm for face recognition under pose and illumination variations is verified using CMU MultiPIE and Extended YaleB face datasets. Different experimental settings and comparisons with existing approaches has been presented below.

4.1 CMU MultiPIE

MultiPIE face dataset collected by Carnegie Mellon University contains 337 subjects captured in four different sessions under 15 view points and 20 illumination conditions (examples are shown in Figure 4.1). The number of subjects across the four sessions is 129 subjects. Face detection and cropping of this subset of data was done using Haar Cascades and manual cropping (due to limitations of face detection algorithms on pose variation). The resultant face images were then resized to 90x90 resolution. In these experiments, 6 different view points -45, -30, 0, 15, 30 and 45 degrees have been considered. Source domain contains poses 0, 45 and -45 degrees, while target domain varies according to experimental setups.

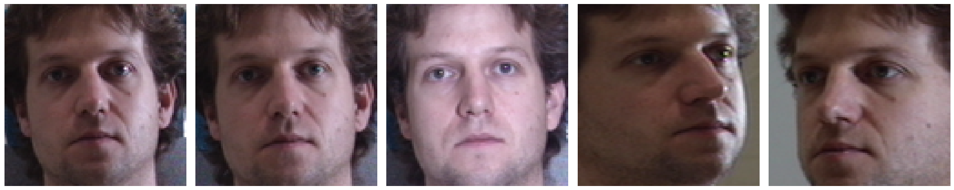

Figure 4.1: Samples of CMU MultiPIE face dataset for pose and illumination variations

4.1.1 Experimental Protocol

4.1.1.1 Experimental Setup-1

For experimental setup-1, the following protocol has been followed: source domain contains instances having pose variations of 0, 45 and -45 degrees. For each user, the source dictionary is learnt on this data. Similarly, target domain contains instances having pose variations of 30, 15 and -30 degrees, thereby, ensuring that the training and testing instances were mutually exclusive. Using the proposed algorithm, the learnt dictionaries are adapted to get the best representation of the target domain. Testing of trained and adapted classifier was done on probe images containing poses at 45, 30, 15, 0, -30 and -45 degrees.

4.1.1.2 Experimental Setup-2

For experimental setup-2, the following protocol has been followed: source domain contains instances from pose variations of 0, 45 and -45 degrees on which the source dictionary is learnt. Target domain consists of images having pose variations of 30, 15 and 45 degrees, individually. The source dictionaries are adapted for the target domains and results were obtained on set of probe images containing 30, 15 and 45 degrees exclusively.

4.1.1.3 Experimental Setup-3

For experimental setup-3, following protocol has been followed: source domain contains instances from poses 0, 45 and -45 degrees on which the source dictionary is learnt. Target domain consists of images having pose variations of 30, 15 and 45 degrees (for both left and right view), individually. The source dictionaries are adapted for the target domains and results were

Table 4.1: Training and testing split for CMU MultiPIE for different experimental setups

	Training		Testing
	Source Domain	Target Domain	Probe Images
Experimental Setup-1	11,610	11,610	15,480
Experimental Setup-2	11,610	3,870	5,160
Experimental Setup-3	11,610	3,870	5,160

obtained on a set of probe images containing 30, 15 and 45 degrees (for both left and right view), exclusively.

4.2 Extended YaleB

The extended YaleB dataset contains 16128 face images of 28 subjects captured under 9 view points and 64 different illumination conditions (examples are shown in Figure 4.2). For our research we considered all the view points and illumination conditions. Face detection and cropping of this dataset was mainly done using Haar Cascade with some manual cropping. The resultant face images were then resized to 60x60 resolution.

4.2.1 Experimental Protocol

The Extended YaleB dataset consists of 28 individuals with a total of nine different views for each subject and 64 different illumination conditions per view. The source domain contains 34 images for each subject, chosen at random, for poses P00, P01, P07 and varying lighting conditions. Class-wise source dictionary is learned on this data. Similarly, target domain contains 34 images of each subject having pose variations mutually exclusive from source domain. The images were cropped and normalized to 60x60. Using the proposed algorithm, the learnt dictionaries are adapted to get the best representation of the target domain. Testing of trained and adapted classifier was done on test data containing 30 images per subject, view and varying lighting conditions.

Figure 4.2: Samples of Extended YaleB face dataset for pose and illumination variations

Table 4.2: Training and testing split for Extended YaleB dataset

	Training		Testing
	Source Domain	Target Domain	Probe Images
Experimental Setup	2,856	5,712	7,560

4.3 Results

Results obtained with the proposed algorithm is evaluated on CMU MultiPIE and extended YaleB dataset using experimental protocol mention in Section 4.1 and 4.2 are summarized as follows:

- In **Experimental setup-1**, where testing was done on probe images containing poses 0, 15, -30, 30, 45 and -45 degrees, a rank-1 accuracy of 73.06% was obtained. This shows that the proposed algorithm can handle the randomness in the target domain with respect to varying pose and illumination while other algorithms aim to adapt the source domain to a specific target domain. Algorithms like GMLDA [27], SDDL [28] and FDDL [33] adapt its source domain to either 30, 15 or 45 degrees individually and do not take into account the case of all poses together.
- Similar results were obtained for extended YaleB dataset where LR-GSRC outperformed other algorithms by obtaining a rank-1 accuracy of 93.2%, as mentioned in Table 4.5
- Adaptive dictionary learning based algorithms like SDDL and LR-GSRC tend to perform well for domain adaptation under pose and illumination variations. LR-GSRC obtained the highest rank-1 accuracy for 15 and 45 degrees (99.8 and 99.5 respectively) while obtaining good accuracy of 98.6 for 30 degrees. This shows that sparse representation and adaptive

dictionary learning framework is a progressive approach towards domain adaptation

- Most of the algorithms in literature can handle either left or right view of a pose at a time while performing domain adaptation. However, introducing group sparsity with existing dictionary learning framework, trains the classifier to learn both left and right view of a pose simultaneously such that, even if target domain contains different views for a class it is able to assign it to the correct identity. Experimental setup-3 verifies this property of the proposed algorithm by obtaining great accuracies on +/- (30,45 and 15) degrees as shown in table 4.2

- The adaptive dictionary learning approach interpolates the path between the source and target domain by learning intermediate dictionaries at each iteration. This builds a linear path from the source to target dictionary such that, given a pose that lies on this linear path, classifier should be able to assign it to the right identity.

Table 4.3: Comparison of proposed algorithm with other Domain Adaptation algorithms on CMU MultiPIE dataset for Experimental Setup-2

Method	15^0	30^0	45^0	Average Accuracy
GMLDA [27]	99.7	99.2	98.6	99.2
GMMFA [27]	99.7	99.0	98.5	99.1
LDA+CCA [26]	95.9	94.9	93.6	94.8
FDDL [33]	96.8	90.6	94.4	93.9
SDDL [28]	98.4	98.2	98.9	98.5
LR-GSRC	**99.8**	98.6	**99.5**	**99.3**

Table 4.4: Performance of proposed algorithm on CMU MultiPIE dataset for Experimental Setup-3

Method	$+/-$ 15^0	$+/-$ 30^0	$+/-$ 45^0	Average Accuracy
LR-GSRC	90.9	83.4	89.1	87.8

Along with Rank-1 accuracies, CMC curves for all four setups have been shown in Figure 4.3, 4.4, 4.5 and 4.6. The comparison between proposed algorithm and exsiting work has also been shown graphically in Figure 4.7.

Table 4.5: Comparison of proposed algorithm with other Domain Adaptation algorithms on extended YaleB dataset

Method	Identification Rate
SRC [31]	90.0
DKSVD [36]	75.3
SVM	88.8
DLSI	85.0
FDDL [33]	91.9
LR-GSRC	**93.2**

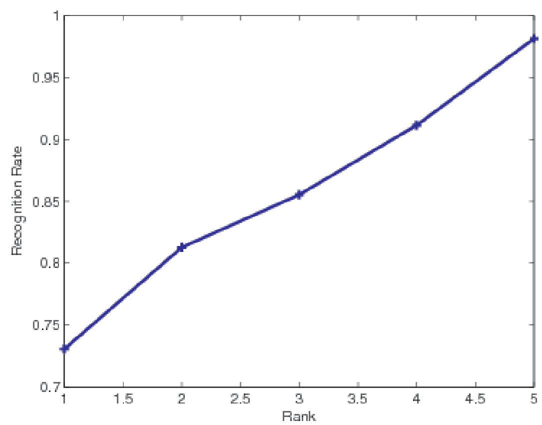

Figure 4.3: Face identification accuracy on CMU MultiPIE for proposed algorithm with multiple poses in target domain (Experimental Setup-1)

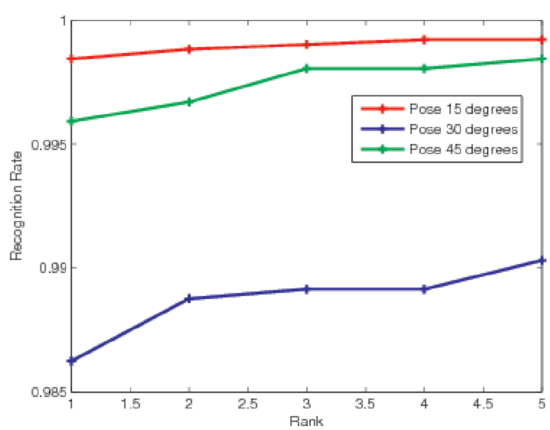

Figure 4.4: Face identification accuracy on CMU MultiPIE for proposed algorithm with each domain individually (Experimental Setup-2)

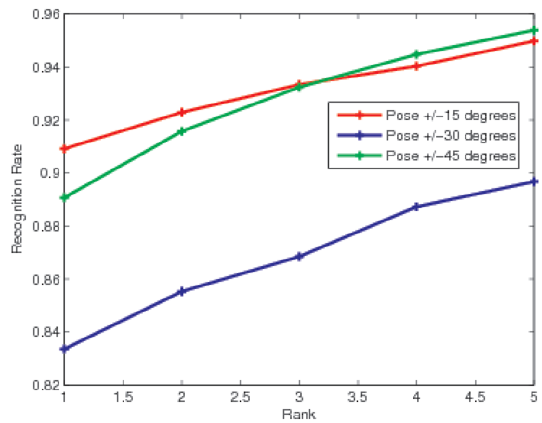

Figure 4.5: Face identification accuracy on CMU MultiPIE for proposed algorithm with each domain individually containing both left and right view (Experimental Setup-3)

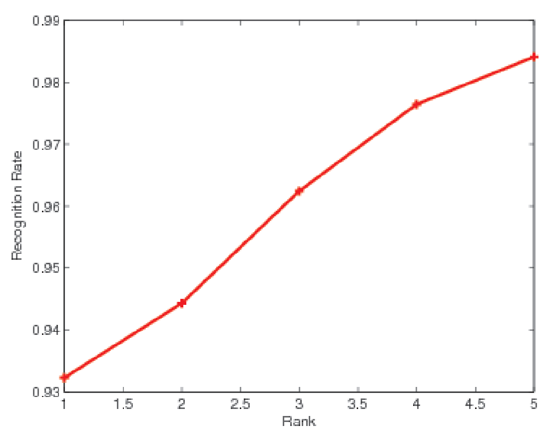

Figure 4.6: Face identification accuracy on extended YaleB face dataset for proposed algorithm

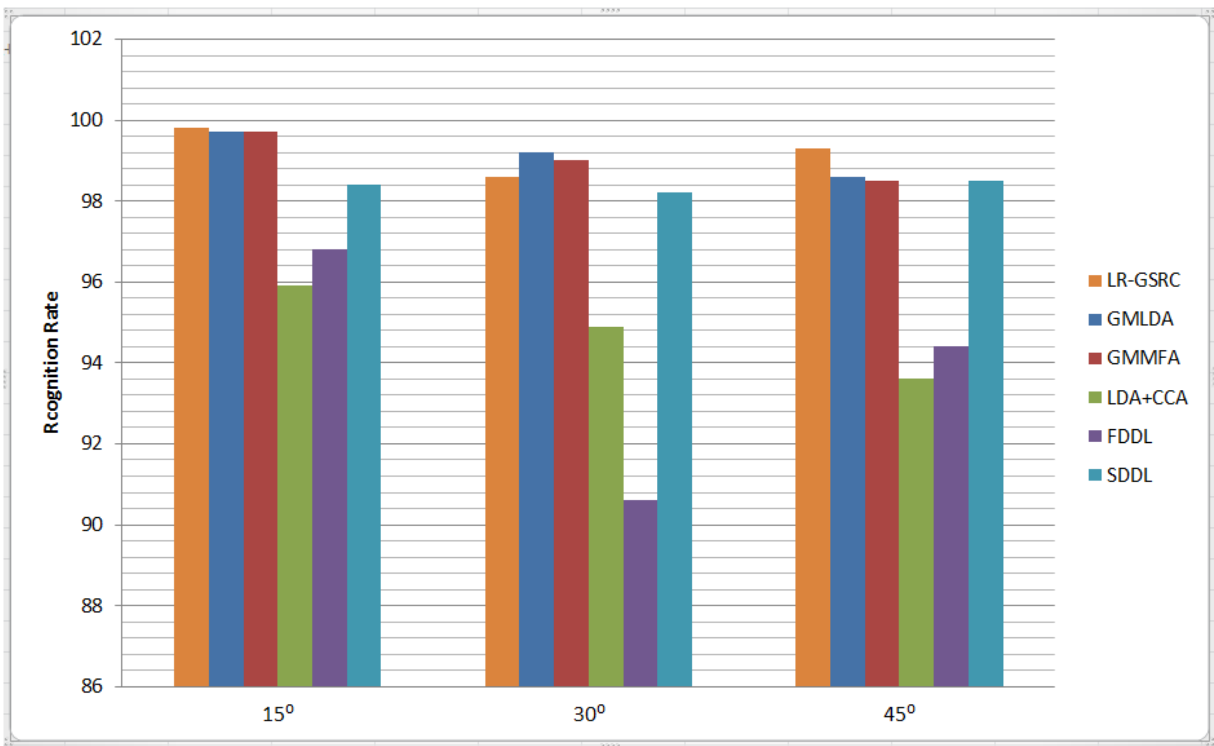

Figure 4.7: Comparison of LR-GSRC on CMU MultiPIE with some exsiting work

Chapter 5

Conclusion and Future Work

In this research, we proposed a novel framework, Low Rank Group Sparse Representation based Classifier, to learn low rank dictionary and Group Sparse coefficients for the problem of face recognition with varying pose and illumination conditions. Results on the CMU MultiPIE and extended YaleB dataset support the effectiveness of the proposed algorithm for domain adaptation. Therefore, there is a scope for extending this work towards various other domains including resolution, blurdness, occlusion, etc. Some literature has shown that adpative dictionary learning framework performs fairly well for object recognition as well [28]. Thus, the scope of this work is not just limited to face recognition and it can be extended for object recognition as well.

In the proposed framework, we have focussed to deal with pose and illumination variations for view point between -45 to 45 degrees. Looking at the results obtained we aspire to expand our view point to the range of -90 to 90 degrees for full coverage of pose and illumination variations. In this case, keeping source domain as 0, 90 and -90 degrees domain adaptation will be done on -70, -60, -45, -30, -15, 0, 15, 30, 45, 60 and 70 degrees. The experiments and results can be obtained on CMU MulitPIE dataset, which provides data according to this specification, with respect to the experimental protocol defined for this dataset.

Also as mentioned above, there is a scope of extending this work beyond pose and illumination covariates in face recognition and diversifying it to object recognition as well. Therefore, this research work can be extended to the following:

- Face recognition across resolution and blurdness: To solve this problem we can take advangtage of group sparcity in our proposed framework such that using resoltuion, blurdness and normal images as three different groups in LR-GSRC we can adapt to learn good representation for target domain. This might help to deal with the problems of blurdness and varying resolutions in our testing data
- Object recognition: The Amazon dataset provides three domains under which images of different objects have been captured. These images are captured using dslr, webcam and images available on Amazon. To approach this problem, we can use two domains as source as two different groups for our LR-GSRC and adapt it to the third domain. Testing and results will be obtained for this third domain

Bibliography

[1] Aharon, M., Elad, M., and Bruckstein, A. K-svd: An algorithm for designing overcomplete dictionaries for sparse representation. *IEEE Transactions on Signal Processing 54*, 11 (2006), 4311–4322.

[2] Blitzer, J., McDonald, R., and Pereira, F. Domain adaptation with structural correspondence learning. In *Proceedings of the 2006 conference on Empirical Methods in Natural Language Processing* (2006), Association for Computational Linguistics, pp. 120–128.

[3] Chai, X., Shan, S., Chen, X., and Gao, W. Locally linear regression for pose-invariant face recognition. *IEEE Transactions on Image Processing 16*, 7 (2007), 1716–1725.

[4] Eldar, Y. C., Kuppinger, P., and Bölcskei, H. Block-sparse signals: Uncertainty relations and efficient recovery. *IEEE Transactions on Signal Processing 58*, 6 (2010), 3042–3054.

[5] Elhamifar, E., and Vidal, R. Sparse subspace clustering. In *IEEE Conference on Computer Vision and Pattern Recognition* (2009), pp. 2790–2797.

[6] Elhamifar, E., and Vidal, R. Robust classification using structured sparse representation. In *IEEE Conference on Computer Vision and Pattern Recognition* (2011), pp. 1873–1879.

[7] Favaro, P., Vidal, R., and Ravichandran, A. A closed form solution to robust subspace estimation and clustering. In *Computer Vision and Pattern Recognition,IEEE Conference* (2011), pp. 1801–1807.

[8] Georghiades, A. S., Belhumeur, P. N., and Kriegman, D. J. From few to many: Illumination cone models for face recognition under variable lighting and pose. *IEEE Transactions on Pattern Analysis and Machine Intelligence 23*, 6 (2001), 643–660.

[9] Gopalan, R., Li, R., and Chellappa, R. Domain adaptation for object recognition: An unsupervised approach. In *IEEE International Conference on Computer Vision* (2011), pp. 999–1006.

[10] Goswami, G., Mittal, P., Majumdar, A., Vatsa, M., and Singh, R. Group sparse representation based classification for multi-feature multimodal biometrics. *Information Fusion 32*.

[11] Gross, R., Matthews, I., Cohn, J., Kanade, T., and Baker, S. Multi-pie. *Image and Vision Computing 28*, 5 (2010), 807–813.

[12] Huang, G. B., Ramesh, M., Berg, T., and Learned-Miller, E. Labeled faces in the wild: A database for studying face recognition in unconstrained environments. Tech. rep., Technical Report 07-49, University of Massachusetts, Amherst, 2007.

[13] Huang, K., and Aviyente, S. Sparse representation for signal classification. In *Advances in Neural Information Processing Systems* (2006), pp. 609–616.

[14] Keshavan, R., Montanari, A., and Oh, S. Matrix completion from noisy entries. In *Advances in Neural Information Processing Systems* (2009), pp. 952–960.

[15] Kulis, B., Saenko, K., and Darrell, T. What you saw is not what you get: Domain adaptation using asymmetric kernel transforms. In *IEEE Conference on Computer Vision and Pattern Recognition* (2011), pp. 1785–1792.

[16] Li, L., Li, S., and Fu, Y. Discriminative dictionary learning with low-rank regularization for face recognition. In *10th IEEE International Conference and Workshops on Automatic Face and Gesture Recognition* (2013), pp. 1–6.

[17] Ma, L., Wang, C., Xiao, B., and Zhou, W. Sparse representation for face recognition based on discriminative low-rank dictionary learning. In *IEEE Conference on Computer Vision and Pattern Recognition* (2012), pp. 2586–2593.

[18] Mairal, J., Bach, F., Ponce, J., Sapiro, G., and Zisserman, A. Discriminative learned dictionaries for local image analysis. In *IEEE Conference on Computer Vision and Pattern Recognition* (2008), pp. 1–8.

[19] Majumdar, A., and Ward, R. K. Robust classifiers for data reduced via random projections. *IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics 40*, 5 (2010), 1359–1371.

[20] Passalis, G., Perakis, P., Theoharis, T., and Kakadiaris, I. A. Using facial symmetry to handle pose variations in real-world 3d face recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence 33*, 10 (2011), 1938–1951.

[21] Pentland, A., Moghaddam, B., and Starner, T. View-based and modular eigenspaces for face recognition. In *IEEE Computer Society Conference on Computer Vision and Pattern Recognition* (1994), pp. 84–91.

[22] Pham, D.-S., and Venkatesh, S. Joint learning and dictionary construction for pattern recognition. In *IEEE Conference on Computer Vision and Pattern Recognition* (2008), pp. 1–8.

[23] Qiu, Q., Ni, J., and Chellappa, R. Dictionary-based domain adaptation methods for the reidentification of faces. In *Person Re-Identification*. Springer, 2014, pp. 269–285.

[24] Saenko, K., Kulis, B., Fritz, M., and Darrell, T. Adapting visual category models to new domains. In *Computer Vision–ECCV 2010*. Springer, 2010, pp. 213–226.

[25] Schroff, F., Kalenichenko, D., and Philbin, J. Facenet: A unified embedding for face recognition and clustering. In *IEEE Conference on Computer Vision and Pattern Recognition* (June 2015).

[26] Sharma, A., Dubey, A., Tripathi, P., and Kumar, V. Pose invariant virtual classifiers from single training image using novel hybrid-eigenfaces. *Neurocomputing 73*, 10 (2010), 1868–1880.

[27] Sharma, A., Kumar, A., Daume III, H., and Jacobs, D. W. Generalized multiview analysis: A discriminative latent space. In *IEEE Conference on Computer Vision and Pattern Recognition* (2012), pp. 2160–2167.

[28] Shekhar, S., Patel, V., Nguyen, H., and Chellappa, R. Generalized domain-adaptive dictionaries. In *IEEE Conference on Computer Vision and Pattern Recognition* (2013), pp. 361–368.

[29] Singh, R., Vatsa, M., Ross, A., and Noore, A. Performance enhancement of 2d face recognition via mosaicing. In *Fourth IEEE Workshop on Automatic Identification Advanced Technologies* (2005), pp. 63–68.

[30] Wang, C., and Mahadevan, S. Manifold alignment without correspondence. In *International Joint Conference on Artificial Intelligence* (2009), vol. 2, p. 3.

[31] Wright, J., Yang, A. Y., Ganesh, A., Sastry, S. S., and Ma, Y. Robust face recognition via sparse representation. *IEEE Transactions on Pattern Analysis and Machine Intelligence 31*, 2 (2009), 210–227.

[32] Yang, J., Yan, R., and Hauptmann, A. G. Cross-domain video concept detection using adaptive svms. In *15th international conference on Multimedia* (2007), pp. 188–197.

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

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

[35] Yuan, X.-T., Liu, X., and Yan, S. Visual classification with multitask joint sparse representation. *Image Processing, IEEE Transactions on 21*, 10 (2012), 4349–4360.

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

[37] Zhang, Y., Jiang, Z., and Davis, L. Learning structured low-rank representations for image classification. In *IEEE Conference on Computer Vision and Pattern Recognition* (2013), pp. 676–683.

[38] Zhu, Z., Luo, P., Wang, X., and Tang, X. Deep learning identity-preserving face space. In *IEEE International Conference on Computer Vision* (2013), pp. 113–120.

[39] Zou, W. W., and Yuen, P. C. Very low resolution face recognition problem. *IEEE Transactions on Image Processing 21*, 1 (2012), 327–340.