2014

A Supervised Low-Rank Matrix Decomposition for Matching

Sajid Sharlemin
West Virginia University

A Supervised Low-Rank Matrix Decomposition for Matching

by

Sajid Sharlemin

Thesis submitted to the
Statler College of Engineering and Mineral Resources
at West Virginia University
in partial fulfillment of the requirements
for the degree of

Master of Science
in
Computer Science

Don Adjeroh, Ph.D.
Xin Li, Ph.D.
Gianfranco Doretto, Ph.D., Chair

Lane Department of Computer Science and Electrical Engineering

Morgantown, West Virginia
2014

Keywords: Low Rank Methods, Face Recognition, Person Re-identification

Abstract

A Supervised Low-Rank Matrix Decomposition for Matching

by

Sajid Sharlemin
Master of Science in Computer Science

West Virginia University

Gianfranco Doretto, Ph.D., Chair

Human identification from images captured in unconstrained scenarios is still an unsolved problem, which finds applications in several areas, ranging from all the settings typical of video surveillance, to robotics, metadata enrichment of social media content, and mobile applications. The most recent approaches rely on techniques such as sparse coding and low-rank matrix decomposition. Those build a generative representation of the data that on the one hand, attempts capturing all the information descriptive of an identity; on the other hand, training and testing are complex to allow those algorithms to be robust against grossly corrupted data, which are typical of unconstrained scenarios.

This thesis introduces a novel low-rank modeling framework for human identification. The approach is supervised, gives up developing a generative representation, and focuses on learning the subspace of nuisance factors, responsible for data corruption. The goal of the model is to learn how to project data onto the orthogonal complement of the nuisance factor subspace, where data become invariant to nuisance factors, thus enabling the use of simple geometry to cope with unwanted corruptions and efficiently do classification. The proposed approach inherently promotes class separation and is computationally efficient, especially at testing time. It has been evaluated for doing face recognition with grossly corrupted training and testing data, obtaining very promising results. The approach has also been challenged with a person re-identification experiment, showing results comparable with the state-of-the-art.

Acknowledgements

I would first like to thank my committee chair and advisor, Dr. Gianfranco Doretto, for giving me the opportunity to work with him and his students. This thesis would not be possible without his constant guidance and support.

I would also like to thank Dr. Donald Adjeroh and Dr. Xin Li for being on my committee. I have been fortunate to have had the opportunity to take courses with all of my committee members, and their teachings have been essential to my understanding of the subject.

Next, I would also like to thank the students in the Computer Vision research lab with whom I've had the pleasure of working alongside. In particular, I would like to thank my colleague Farzad Siyahjani , who has been a great help to me.

Finally, I would like to express my gratitude to my family.

Contents

List of Figures

List of Tables

Notation

We use the following notation and symbols throughout this thesis.

$\|\cdot\|_*$	:	Nuclear Norm
$\|\cdot\|_F$	:	Frobenius Norm
$\mathcal{V}$	:	Variation Subspace
$\mathcal{B}$	:	Invariant Subspace
$P_{(\cdot)}$	:	Projection Operator
$(\cdot)^\top$	:	Transpose
α	:	penalty weights
β	:	penalty weights
γ	:	penalty weights
X	:	Input Data
A	:	Low-Rank Component
B	:	Invariant Component
E	:	Sparse Error Component

Chapter 1

Introduction

Recent research by the Computer Vision community has proven that sparse signal representation is an extremely powerful way to represent high dimensional image data. In most cases an input signal is represented as a linear combination of a few items from a set or dictionary D. Using sparse representation techniques researchers have achieved significant results on image classification [1, 2, 3]. For classification methods based on sparse representation the dictionary size and quality plays a vital role. Sparse coding algorithms tend to suffer efficiency when the dictionary size is too large. Well constructed compact dictionary can make sparse coding based image classification algorithms computationally efficient. However performance of these methods deteriorates drastically in the presence of large error contamination like occlusion, lighting variations and too much noise were present in the training dataset.

Low rank representation is an effective method for doing subspace clustering. The main goal of low rank matrix recovery methods is to determine a low rank matrix approximation from corrupted input datasets. These methods have been used successfully for object detection, segmentation, tracking, background subtraction and even image classification. Sparse representation based classification has shown robustness to high degree of noise and occlusion in test images. However, the method is sensitive when learning a dictionary from training samples corrupted from nuisance factors like occlusion, lighting variation and so on. To tackle this problem, low rank matrix decomposition algorithms have been developed which can learn a representational dictionary even with input data highly affected by nuisance

factors.

Human identification is one of the major challenges in Computer Vision. Human identification is important for security purposes and biometrics. In unconstrained scenarios identification of a human subject can be done with face images or with full body images. In this kind of real life scenario, images that will be used for testing or training is expected to be severely contaminated by nuisance factors. Low rank methods have been used successfully for face recognition. However, till now these methods haven't been applied to more complicated scenarios like person re-identification, where there is a lot of pose and illumination variances to cope with.

Current low rank matrix decomposition algorithms developed for face recognition make the assumption that the images representing one identify lie in a low dimensional space which is a subspace of the column spaces spanned from all the images of that identity, and model the corruption noise as a sparse matrix. We have used a low rank matrix decomposition in a different way to model the contaminated data. Rather than learning a subspace spanned by each identity from the data with nuisance factors we try to learn the nuisance subspace that caused corruption of the data and its orthogonal complement subspace. For learning this decomposition we use the Augmented Lagrange Multiplier method. We have used our proposed low rank matrix decomposition for face recognition on a publicly available data sets and obtained very promising results. We also tried our method on some public person re-identification data sets which is a very challenging scenario considering that in this case there is a large amount of image misalignment, compared to face recognition where low rank methods have been used very successfully. We did find some encouraging results on person re-identification data sets, which are comparable to the state-of-the art results.

Chapter 2

Literature Review

Robust PCA has been introduced to solve the shortcomings of classical PCA which mails to find the right underlying distribution under the presence of gross noise or outliers. A convex optimization problem needs to be solved to solve RPCA. This optimization can be treated as a general convex optimization problem and solved by any off-the-shelf interior point solver (e.g, CVX [4]), after being formulated as a semi definite program [5]. Although interior point methods usually takes few iterations to converge, they have difficulty in handling large matrices because of the complexity of computing step direction is $\mathrm{O}\big(m^6\big)$, where m is the dimension of the matrix. Due to this computational inefficiency generic interior point solvers cannot handle matrices with dimensions larger than $m = 10^2$. However in Computer Vision problems it is not uncommon to find matrices with dimension $m = 10^4$ to 10^5; and applications in web search and bioinformatics can easily involve matrices of dimension $m = 10^6$ to 10^7. So interior point method solvers are too limited for Robust PCA to be practical for many real world applications.

The interior point solvers do not scale well because they rely on second-order information of the objective function. To overcome the scalability issue we need to use only first-order information and make full advantage of the special properties of this class of convex optimization problems.For example , it has been recently shown that the first-order iterative thresholding(IT) algorithms can be very efficient for ℓ_1-norm minimization problems arising in compressed sensing [6, 7, 8]. It has also be shown in [9, 10] that the same techniques can be used to minimize the nuclear norm for the matrix completion (MC) problem, namely

recovering a low-rank matrix from an incomplete but clean subset of its entries [11].

As Robust PCA problem involves minimizing a combination of both the ℓ_1-norm and the nuclear norm in [12] the authors have adopted the iterative thresholding technique to solve RPCA and obtained similar convergence and scalability properties. However the proposed iterative thresholding technique to solve RPCA converges extremely slowly. Typically it needs about 10^4 iterations to converge while each iteration cost as much as one SVD. Due to this slow convergence even for matrices with dimension $m = 800$ the algorithm takes around 8 hour on a typical PC. To alleviate the slow convergence of the iterative thresholding method [12] two new algorithms were proposed in [13] for solving the optimization problem of RPCA, which in some sense complementary to each other. The first one is an accelerated proximal gradient(APG) algorithm applied to the primal, the second one is a gradient-ascent algorithm applied of the dual of the original problem of RPCA . It has shown from simulations that both of these methods are at least 50 times faster than the iterative thresholding method [13].

In the paper [14] authors presented an algorithm for matrix decomposition that utilize techniques of augmented Lagrangian multipliers (ALM). The exact ALM (EALM) method proposed is shown to have a Q-linear convergence speed, while the APG is in theory only sub-linear. A slight improvement over the exact ALM (EALM) leads to the inexact ALM (IALM) method which converges practically as fast as the exact ALM, however the required partial SVDs is significantly lower. Simulations show that IALM is at least *five times faster* than APG, and its precision is alos higher. In particular, the number of non-zeros in E computed by IALM is much more accurate than the APG, which tends to leave many small non-zero terms in E.

Low-rank methods have been applied and extended in many image processing and Computer Vision applications with promising outcome [1, 15, 16, 17, 18, 19]. There are some interesting works that show how the combination of low-rank method and sparse modeling of the pixels in an image with parametric transformations of the image domain can be used for holistic symmetry detection and rectification, In [20] the authors used lo-rank matrix decomposition to learn intrinsic invariant low-rank texture of objects and extract linear transformation of the 3-D scene over associated planar region. Using this new proposed method

the authors were able to overcome the lack of invariant image features under protective transformation, where it enables to recover the exact low-rank structure by simultaneously discarding sparse noise and transforming the feature space.The resulting algorithm, called transformation-invariant low-rank texture (TILT), has been successfully applied to practical problems such as urban 3D reconstruction, calibration and optical character recognition. In [21] TILT has been extended to deal with low-rank textures on generalized cylindrical surfaces in 3D space and thus is naturally robust to occlusion and other corruptions.

In photo metric stereo problem the goal is to estimate the normal map of a scene given multiple 2D images taken under the same viewpoint but different lighting conditions. The authors in [22] used recent ideas from the theory of sparse and low-rank matrix decomposition formulated a convex optimization problem and were able to recover the low-rank structure from the input matrix despite the presence of large, sparse errors. Once the low-rank matrix is recovered, the normal map was easily computed.

Low-rank method has also been applied to pixel wise image alignment where the problem is to align multiple images of an object to a fixed canonical template in [23] because the images of same object if represented as a matrix where each image is a column vector, the matrix is expected to have low-rank.

[24] introduced low-rank subspace clustering (LRSC) in a convex formulation using the idea of adding the self expressiveness constraint into low rank decomposition. They were able to decompose subspace clusters in an unsupervised manner from noisy data. [25] added a fixed rank constraint to the problem of low-rank representation and showed that it can be used for feature extraction. [26] has described theory and applications for sparse subspace clustering. [27] introduced a dimension reduction method where it learns a projection matrix and a sparse coefficients matrix to jointly reduce the dimensions of the data points and does the unsupervised clustering in a single objective function. [24] used an structured matrix to do supervised clustering of data points, where with a new goal function the authors decomposed the data points to low-rank and sparse noise where the low-rank matrix is the linear combination of the basis in a dictionary and using a structured matrix they force subjects from the same class lie on the same subspace.

Low-rank matrix decomposition has also been used in face recognition recently and very

promising results has been obtained. Earlier it was common to apply existing techniques such as Eigenfaces [28], Fisherfaces [29], or Laplacian faces [30] to reduce dimension of face images. As a result the derived subspace was expected to achieve improved recognition performance. These methods however are not robust to outliers or gross noise. Recently Robust PCA has been proposed to alleviate these shortcomings [31, 32, 33].

Sparse representation based classification (SRC) [2] has shown very promising results on face recognition. It considers each test image as a sparse linear combination of the training samples by solving an l^1 minimization problem. If the test image is corrupted, SRC exhibits robustness to face occlusion and corruption. As SRC requires the training images to be well aligned for reconstruction purposes, in [34] the authors further extends it to deal with face misalignment and illumination variations. Yang et. al [3, 35] also modified SRC based method to tackle occlusions in face images. One shorting of the method is that it might not generalize well if both training and test images are corrupted. In [1] the authors addressed the problem of robust face recognition by low-rank approximation with structural incoherence where both training and test image data are corrupted and there is no prior knowledge about the type of corruption. This method has very large computational cost in the presence of large number of classes as denoising is done class by class. [36] enhances a sparse coding dictionary's by learning a low-rank sub-dictionary for each class. This method is time consuming and might increase the redundancy in each sub-dictionary thus not guaranteeing consistency of sparse codes for signals from same class. [17] presents an image classification framework by using non-negative sparse-coding, low-rank and sparse matrix decomposition. A linear SVM classifier is used for the final classification. In [18] the authors effectively constructs a reconstructive and discriminative dictionary. Based on this dictionary structured low-rank and sparse representations are learned for classification.

In Computer-Vision research community there has been increasing interest in matching people across disjoint camera views in a multi-camera system, commonly known as person re-identification problem [37, 38, 39, 40]. Most existing studies have tried to solve person re-identification problem by seeking a more distinctive and stable representation of people's appearance ranging widely from color histogram [37, 39], graph model [41], spatial co-occurance [40] representaion model [40], principal axis histogram [38], rectangle region

histogram [42] to combination of multiple features [39, 43]. After the feature extraction stage many existing methods simply choose a standard distance measure such as l_1 norm [40], l_2 norm based distance [38] or Bhattacharyya distance [39]. Unfortunately under severe change of viewing condition such as different pose, occlusion and illumination computing a set of distinctive and stable feature is extremely hard if not impossible in realistic scenarios. Machine learning approaches for solving perosn re-identification has been proposed for salient feature learning [44], attributes [45] and ranking functions [46]. Metric learning methods are more recent approaches and include methods like Large Margin Nearest Neighbors (LMNN) [47], Metric Learning by Collasping Classes [48], Probabilistic Relative Distance Comparison [49] , Pairwise Constrained Component Analysis (PCCA) [50]. There is not any literature on applying low-rank method to solve person re-identification problem. This is a much more difficult scenario compared to face recognition as the degree of alignment is typically much less compared to the face recognition problem settings. In our experiments we challenge our approach by comparing it against one metric learning method, despite the fact that we have not exploited the possibility to use a dedicated descriptor, or to do an extensive learning with a dedicated dataset and our approach has to deal to with the huge amount of misalignment.

Chapter 3

Robust Principal Component Analysis

Principal Component Analysis (PCA) is a statistical technique that is widely used for applications such as dimensionality reduction, lossy data compression, feature extraction and data visualization.It is also known as it is also named the discrete Karhunen-Loève transform (KLT) in signal processing. Among many destinations two are commonly used that boil down to the same algorithm. PCA can be defined as the orthogonal projection of the data onto a lower dimensional linear space, known as the *principal subspace*, such that the variance of the projected data is maximized [51]. It can also be defined as the linear projection that minimizes the average projection cost, defined as the mean squared distance between the data points and their projections. Basically PCA uses an orthogonal transformation to convert a set of possibly correlated variables into a set of values that are called principal components. The number of principal components is less than number of original variables. The transformation is such that the first principal component has the largest possible variance (that is, accounts for as much of the variability as possible) and each succeeding components in turn has the highest possible variance under the constraint that it is orthogonal i.e , uncorrelated with the previous components. The subspace that is spanned by all the principal component make the principal subspace.

Let us consider a data set of observation $X = [x_1, x_2, \ldots, x_n]$ where x_m is a column vector with dimensionality m. PCA assumes that the given data was generated by perturbing a matrix $A \in \mathbb{R}^{m \times n}$ whose columns lie on a subspace of dimension $r << m$. The goal of PCA is to project the data onto a space with lower dimension $r << m$ while maximizing the

variance of the projected data. In other words, $X = A + N$ where A is the low rank matrix with rank r and N is the matrix whose entries are i.i.d Gaussian random variables. In this setting, PCA seeks an optimal estimate of A via the following constrained optimization.

$$\begin{aligned} &\text{minimize } \|X - A\| \\ &\text{subject to rank } (A) \leq \text{ rank } (X) \, . \end{aligned} \tag{3.1}$$

We assume for the moment that we know the value of r. PCA involves finding the mean $\bar{X}$ and the covariance matrix S and then finding r eigenvectors $u_1, u_2, \ldots, u_m$ of S corresponding to the m largest eigenvalues $\lambda_1, \lambda_2, \ldots, \lambda_m$. Then we project the original data of dimension m to the principal subspace spanned by the r eigenvectors where $r << m$ to obtain the reduced dimensional data.

The possibility of efficient computation along with optimality properties in the presence of Gaussian noise has made PCA as one of the most popular algorithms used in data analysis, compression among other things. However in today's world data is more often than not corrupted by large errors or can even be incomplete. In the presence of large noise corrupting some of the data PCA estimate can be far from the underlying true distribution of the data owing to the fact that classical PCA is very sensitive to gross errors and there is no theoretical framework to deal with incomplete missing data. In fact even if only one entry of X is arbitrarily corrupted then estimated A obtained by classical PCA can be far from the true A we are seeking. Another adversary is that rank m of the data matrix X needs to known beforehand, which is seldom the case in real world problems.

Unfortunately gross errors are not very uncommon in modern applications such as image processing, web data analysis, bioinformatics, where some measurements can be arbitrarily corrupted due to occlusions, malicious tampering, sensor failure or simply irrelevant to the low rank structure we seek to identify. A number of natural approaches to robustifying PCA have been explored in the literature over the decades. The representative approaches include influence function techniques, multivariate trimming, alternating minimization and random sampling technique. Unfortunately none of these techniques yield polynomial time algorithm with strong performance guarantee under broad conditions. The new problem that we are

considering here can be considered as an idealized version of *Robust* PCA in which we aim to recover a low-rank matrix A from highly corrupted measurements $X = A + E$. Unlike small noise term N in classical PCA the entries in E can have arbitrarily high magnitude.

At first sight, the separation problem seems really hard to solve since the number of unknowns to infer A and E is twice as many as the given data input $X \in \mathbb{R}$. Furthermore it looks even more challenging that we expect to reliably obtain the low-rank matrix A with errors in E of arbitrarily large magnitude.

However in RPCA paper the authors showed that this problem can not only be solved but it can be solved by tractable convex optimization. Let $\|D\|_* = \sum_i \sigma_i(D)$ denote the nuclear norm of matrix X ,i.e, the sum of the singular values of X and let $\|X\|_1 = \sum_{ij} |X_{ij}|$ denote the l_1 norm of X seen as a long vector in $\mathbb{R}$. The authors showed that under rather weak assumptions, the **Principal Component Pursuit** (PCP) estimate solving

$$\begin{aligned} &\text{minimize } \|A\|_* + \lambda \|E\|_1 \\ &\text{subject to } A + E = X \ . \end{aligned} \tag{3.2}$$

(3.2) exactly recovers the low-rank A and the sparse E. Theoretically this is shown to work even if the rank of A grows almost linearly in the dimension of the matrix and the errors in E are up to a constant fraction of all entries.Algorithmically, it can be shown that the above problem can be solved by efficient and scalable algorithms at a cost not too much higher than the cost of solving the classical PCA.

It is not always that the matrix decomposition works. There are instances when it is not possible to make the decomposition work. If the matrix X is both sparse and low-rank then it is hard to decide whether it is a low-rank matrix or a sparse matrix. So to make the problem of robust PCA meaningful we do need to impose that the low-rank component A is not sparse. Another identifiability issue arises if the sparse matrix has low-rank. This is the case if suppose all the non zero entries of E occur in a column or in a few columns only. Suppose for instance, that the first column of E is the opposite of that of A and that all other columns of E vanishes. Then it is certain that we would not be able to recover A and E by any method whatsoever since $X = A + E$ would have a column space equal to, or

included in that of A. To avoid such meaningless scenarios we make the assumptions that the sparsity pattern of the sparse component is selected uniformly at random.

Chapter 4

Model

4.1 Invariant Subspace Representation

We assume that a data point $x \in \mathbb{R}^m$, representing an entity (e.g., the vectorized version of the image pixels of a face), can be modeled by two additive components. The first one, $s \in \mathbb{R}^m$, represents all the information necessary to recognize the entity (e.g., everything that describes the specific identity of the individual depicted by the face image). From a statistical point of view, we can imagine s to be the equivalent of a sufficient statistic for recognition, and we refer to it as the *sufficient component*. The second component, $v \in \mathbb{R}^m$, is meant to represent how the data of a generic entity might change by the effect of nuisance factors, which are not descriptive of any specific entity. For instance, the image of a face might be modified by different lighting conditions, facial expressions, occlusions, etc. It is assumed that all the changes inducible by nuisance factors form a *variation subspace* $\mathcal{V}$, where the *variation component* v is defined. Therefore, a data point is modeled as

$$x \doteq s + v \ . \tag{4.1}$$

If $P_{\mathcal{V}} : \mathbb{R}^m \to \mathcal{V}$ is the projection operator mapping an m-dimensional vector onto $\mathcal{V}$, x can be further decomposed as $x = (P_{\mathcal{V}}s + v) + (s - P_{\mathcal{V}}s)$. In particular, the first component $a \doteq P_{\mathcal{V}}s + v$, is defined in $\mathcal{V}$, whereas the second component $b \doteq s - P_{\mathcal{V}}s$, is defined in the orthogonal complement of the variation space, $\mathcal{V}^{\perp}$.

The decomposition $x = a + b$ has the following property. Let us assume that x_1 and x_2

are two different points representing the same entity. According to (4.1), it must be that $x_1 = s + v_1$ and $x_2 = s + v_2$, because they have been affected by different nuisance factors. This means that $a_1 = P_{\mathcal{V}} s + v_1$, and $a_2 = P_{\mathcal{V}} s + v_2$; however, $b_1 = s - P_{\mathcal{V}} s = b_2$, which highlights that the component b is *invariant* to the changes induced by the nuisance factors. We refer to the subspace where b is defined as the *invariant subspace* $\mathcal{B}$, which will be a subspace of $\mathcal{V}^{\perp}$.

4.2 Recognition Based on the Invariant Subspace

We assume that a set of n training data samples from N different entities, or object classes (e.g. images of people faces, or people whole body appearances), are given, where each class i has n_i samples. Every sample x_j is modeled according to (4.1), and we concatenate the data into a matrix $X = [x_1, x_2, \cdots, x_N] \in \mathbb{R}^{m \times n}$, where $x_i \in \mathbb{R}^{m \times n_i}$ is the training data matrix obtained by lining up the samples for class i.

Model (4.1) has been implicitly adopted by the most successful recent approaches to the face recognition problem. In particular, the SRC method [2] aims at "carefully" composing each of the X_i's in such a way that the selected samples are able to represent the salient components s_i's in the best possible way. The matching between a test point $x = s + v$, and a salient component s_i (i.e. the classification), is based on sparse coding and residual computation, and has demonstrated a remarkable robustness against the variation component v, leading to high recognition rates. The SRC approach has been further improved against potential corruptions of the test data point. For instance, [3] improves upon occlusions and computational cost, [35] robustifies the sparse coding problem by computing a sparsity-constrained maximum likelihood solution, [34] simultaneously handles the misalignment, pose and illumination invariance, and [52] addresses the problem of reducing the large amount of training data needed by SRC to be effective.

To address the more general case where also the training data is highly affected by nuisance factors, and a "careful" composition of X is not possible, the SRC approach has been augmented in different ways. In [1] a low-rank matrix recovery [33] approach is designed for pre-processing the corrupted training data. After this step, the SRC method can be applied

more effectively. Another approach, [53], proposes to apply sparse coding for modeling the sufficient component by learning a dictionary of prototypes, each of which, given by the average of the data in X_i, is meant to approximate s_i. In addition, sparse coding is also used for modeling the variation subspace. The concatenation of the prototype and the variation dictionaries form a new dictionary with which the SRC method can be applied more effectively.

In this work we propose to address the recognition problem with highly corrupted training and testing data by exploiting model (4.1) in a very different way than previous work. The idea is based on a simple observation. Suppose that the projection operator $P_{\mathcal{V}}$ was available. Then, a test sample x could be processed by computing $x - P_{\mathcal{V}}x = b$. Similarly, for the training dataset, following the property of the invariant subspace, computing $X - P_{\mathcal{V}}X$ produces $[b_1 1_{n_1}^\top, b_2 1_{n_2}^\top, \cdots, b_N 1_{n_N}^\top]$, where b_i is the invariant of class i, and 1_{n_i} is a column vector of ones with length n_i. Therefore, recognition could be done by a simple matching between b and the set of b_i's. This means that corruption (or intra-class variability) in training and testing data, as well as recognition could be handled in a very easy, and efficient way with simple geometry tools.

One major challenge of the proposed approach is posed by the case when two different sufficient components $s_1 \neq s_2$, are such that $s_1 - P_{\mathcal{V}}s_1 = s_2 - P_{\mathcal{V}}s_2$. This means it would be impossible to discriminate between the corresponding classes. The supervised learning approach introduced in the following sections will: (1) allow to learn the invariant subspace, and (2) inherently address the challenge just outlined by promoting a uniform inter-class separability.

4.3 Robust PCA and Low-Rank Matrix Recovery

Principal Component Analysis is arguably the most widely used statistical method for data analysis and dimensionality reduction. However it is not without its weakness. PCA has been shown to be sensitive to grossly corrupted input data. Unfortunately gross errors are very common in many modern applications like image processing , web data analysis where some measurements may be arbitrarily corrupted or may be irrelevant to the low

rank structure we wish to identify. To make PCA robust and to mitigate the effect of sparse noise, a lot of approaches have been proposed in the literature , including the introduction of influence functions, alternating minimizing techniques, and low-rank matrix recovery. Low-rank method has been shown to be very efficient and can be solved in polynomial time.

Low-rank matrix recovery seeks to decompose a data matrix X into $A + E$, where A has rank which is much lower than rank of X and E is the associated sparse error. To be precise, given the input data X, low-rank method minimizes the the rank of the matrix A while reducing $\|E\|_0$ to derive the low-rank approximation of X. Since the aforementioned optimization problem is NP-hard, proposed to relax the original problem into the following tractable formulation.

$$\min_{A,E} \|A\|_* + \alpha\|E\|_1 \qquad \text{s.t. } X = A + E \,. \tag{4.2}$$

In (4.2), the nuclear norm $\|A\|_*$ (i.e. the sum of the singular values) approximates the rank of A, and the ℓ_0-norm $\|E\|_0$ is replaced by the ℓ_1-norm $\|E\|_1$, which sums up the absolute values of the entries of E. It is shown in [33] that solving the relaxed version of the problem (4.2) is equivalent to solving the original low-rank matrix approximation problem, as long as the rank of A to be recovered is not too large and the number of errors in E is small (sparse). To solve the optimization problem (4.2) it is possible to apply the efficient method of augmented Lagrangian multipliers (ALM).

4.4 Face recognition by low-rank matrix recovery

For real-world face recognition problems it is expected that training data cannot always be acquired under a controlled settings. Besides illumination, pose, expression variations face image taken by the camera can vary due to the presence of sun-glass, scarf, mask or any such garments. When a image like this is used training the learned face data might over fit the extreme variations in the face images and not model the face of the subjects and thus degrade the performance of recognition process.

Standard LR method processes original data X and produces a low rank matrix A for bet-

ter representation with sparse noise removed. In incoherence paper the authors have argued that face images from different subjects typically share common features (e.g the location of eyes, noses, mouth etc.) and as a result the derived matrix A might not be discriminating. So they proposed to promote the incoherence between low-rank matrices. Introducing such incoherence would prefer the low-rank matrices to be independent as possible. As a result commonly shared features among classes will be suppressed and the discriminating features will be preserved. The authors in the paper added a regularization term to the object function to 4.2 enforce the incoherence between low-rank matrices.

$$\min_{A_i,E_i} \sum_{i=1}^{N} \{\|A_i\|_* + \alpha\|E_i\|_1\} + \eta \sum_{j\neq i} |A_j^T A_i\|_F^2 \qquad \text{s.t. } X_i = A_i + E_i \,. \tag{4.3}$$

In 4.3, the first term performs standard low-rank decomposition for data matrix X. The new term that is added here sums up the Frobenius term between each pair of the low-rank matrices A_i and A_j, which is penalized by the η which balances the low-rank approximation and matrix incoherence. The authors have referred to 4.3 to as low-rank matrix recovery with structural incoherence, aiming to provide improved discrimination ability to the original low- rank decomposition method.

In a newer approach the authors have taken a different route at solving the problem. They argued that [1] performs low-rank recovery using class by class during training which is computationally expensive considering the presence of a large number of classes. They effectively construct a reconstructive and discriminating dictionary from corrupted training data. They show that an efficient representation can be obtained with respect to well-structured dictionary. Associating label information in the training process a discriminate dictionary can be learned from all training samples simultaneously. The learned dictionary encourages images from same subject to lie in the same low-dimensional subspace while different classes lie on different low-dimensional subspace. This boosts the classification result significantly.

For Face recognition the data set consists of many subjects and images of one subject tends to be drawn from the same subspace while samples of different subjects are drawn from different subspace. [24] proves that there is a lowest-rank representation that reveals

the membership of samples. They formulate the low-rank decomposition as follows

$$\min_{Z,E} |Z\|_* + \alpha\|E_i\|_{2,1} \qquad \text{s.t. } X = DZ + E \,. \tag{4.4}$$

Here the data matrix $X = [X_1, X_2, \cdots, X_N]$ contains images of N classes where X_i corresponds to class i. E is the sparse noise component With respect to semantic dictionary D, the optimal representation matrix Z for X should be block diagonal.

$$Z = \begin{pmatrix} Z_1^* & 0 & \cdots & 0 \\ 0 & Z_2^* & 0 & \vdots \\ \vdots & 0 & \ddots & 0 \\ 0 & \cdots & 0 & Z_N^* \end{pmatrix}$$

Here the low-rank reveals the structure information and sparsity identifies which class an image belongs to. Given a dictionary D , the objective function can be written as,

$$\min_{Z,E} |Z\|_* + \alpha\|E\|_1 + \beta\|Z\|_1 \qquad \text{s.t. } X = DZ + E \,. \tag{4.5}$$

Here α and β controls the sparsity of the noise matrix E and the representation matrix Z. The dictionary $D = [D_1, D_2, \cdots, D_N]$ contains N sub dictionaries where D_i corresponds to class i. Let $Z_i = [Z_{i,1}, Z_{i,2}, \cdots, Z_{i,N}]$ be representative of X_i with D being the dictionary. Then $Z_{i,j}$ denotes coefficient s for D_j. For low-rank and sparse representation of data D_i should ideally to exclusive to each subject i. This means that different classes will have different representations. Also every class i should be well represented by its sub dictionary such that $X_i = D_i Z_{i,1} + E_i.Z_i$, where the coefficients for $D_j (i \neq j)$ are nearly all zero.

Q is said to an ideal representation if $Q = [q_1, q_2, \cdots, q_T] \in \mathbb{R}^{K \times T}$ where q_i the code for sample X_i, is of the form of $[0 \cdots 1, 1, 1, , \cdots]^t \in \mathbb{R}^K$ where K is the size of the dictionary and T is the total number of samples. If x_i belongs to class C then the coefficients of q_i for D_C are all 1's where the remaining entries are all 0. Although this decomposition might not result in minimal reconstruction error, low-rank and sparse.

With these formulations the authors propose to learn a semantic structured dictionary by supervised learning. Based on the label information we construct Q in block diagonal fashion for training data. A regulation term is added in the form of $\|Z - Q\|_F^2$ to include

structure information in the dictionary learning process. A dictionary that encourages Z to be close to Q is preferred. The objective function to learn the dictionary is formulated as follows

$$\min_{Z,E,D} |Z\|_* + \alpha\|E\|_1 + \beta\|Z\|_1 + \gamma\|Z - Q\|_F^2 \qquad \text{s.t. } X = DZ + E \,. \tag{4.6}$$

Chapter 5

Model Formulation and Supervised-Learning

5.1 Invariant Subspace Learning

We look at the problem from a completely different perspective to solve the problem of identity recognition using low-rank decomposition.

We begin by observing that since every data point is modeled as $x_j = a_j + b_j$, the training data set X, can be decomposed by $X \doteq A + B$, where $A \in \mathbb{R}^{m \times n}$ collects all the a_j's, and $B \in \mathbb{R}^{m \times n}$ collects all the invariant components, b_j's. We assume that the variation subspace $\mathcal{V}$ has a finite dimension, which is lower than $\min\{m, n\}$. This is reasonable because it states that there are enough data for learning the variation subspace of interest, it allows avoiding overfitting, and it makes the problem tractable. Therefore, attempting to recover A, which in turn allows recovering B, entails solving a low-rank matrix recovery problem.

In practice, the training data will also be affected by noise. Rather than modeling small Gaussian deviations, we admit that a small percentage of the entries of X are corrupted by values not modeled by the variation and invariant components, which means that such noise should be sparse. This will account for data deviations unlikely to be captured by a finite dimensional linear subspace, such as those induced by image saturations, like image glare, or the presence of strong edges. Therefore, if $E \in \mathbb{R}^{m \times n}$ is the matrix of sparse noise, the

model for the training dataset is given by

$$X \doteq A + B + E \ . \tag{5.1}$$

Contrary to previous work we do not attempt to learn a dictionary , and the columns of the low-rank matrix A are meant to span the variation subspace $\mathcal{V}$ not the space of the sufficient components. Discriminability comes from learning the invariant components B, which leads to a very simple rule for classification and can promote class separation approach described next.

To learn model (5.1), standard LR (4.2) is insufficient because we also need to learn the invariant components B. To do so, we need to take into account the geometric, and invariance constraints of (5.1).

5.1.1 Geometric constraint.

In particular, the invariant subspace should be included in the orthogonal complement of the variation subspace $\mathcal{V}^{\perp}$. Therefore, A and B should satisfy the relationship

$$B^{\top} A = 0 \ . \tag{5.2}$$

5.1.2 Invariance constraint.

In addition, given two data points $x_1 = a_1 + b_1 + e_1$ and $x_2 = a_2 + b_2 + e_2$, if they are representative of the same class i, the invariant components should be the same, i.e. $b_1 = b_2$. To express this in an algebraic form, b_1 and b_2 should be the solution to the linear system given by the equations $b_1 = \frac{1}{2}(b_1 + b_2)$, and $b_2 = \frac{1}{2}(b_1 + b_2)$. For n data points, where $B = [B_1, B_2, \cdots, B_N]$, the constraint on the invariant components would be $b_1 = b_2 \cdots = b_{n_1}$, for B_1, $\cdots$, and $b_{n-n_N+1} = b_{n-n_N+2} = \cdots = b_n$, for B_N. This can still be expressed in an algebraic form, by generalizing the system of two linear equations to the following expression.

$$B(I - Q) = 0 \ , \tag{5.3}$$

where I is the identity matrix, and Q is a block-diagonal matrix, given by

$$Q = \begin{pmatrix} \frac{1}{n_1}1_{n_1}1_{n_1}^\top & 0 & \cdots & 0 \\ 0 & \frac{1}{n_2}1_{n_2}1_{n_2}^\top & 0 & 0 \\ \vdots & 0 & \ddots & 0 \\ 0 & \cdots & 0 & \frac{1}{n_N}1_{n_N}1_{n_N}^\top \end{pmatrix}$$

In order to learn A and B, we propose to augment problem (4.2) with model (5.1), the *geometric constraint* (5.2), and the *invariance* (5.3). In particular, to make the problem more tractable, the geometric and invariance constraints are relaxed to the penalty terms $\|B^\top A\|_F^2$, and $\|B(I-Q)\|_F^2$ in the following optimization problem

$$\min_{A,B,E} \|A\|_* + \alpha\|E\|_1 + \beta\|B(I-Q)\|_F^2 + \gamma\|B^\top A\|_F^2 \quad \text{s.t. } X = A + B + E\,, \tag{5.4}$$

where $\|\cdot\|_F$ indicates the Frobenius norm, and α, β, and γ are penalty weights. Note that the addition of the invariance constraint (5.3) as a penalty, through Q injects the training dataset labeling information inside the learning problem, turning it into a supervised approach.

5.1.3 Optimization

In order to solve problem (5.4), we use the exact ALM method [14], and start by computing the augmented Lagrangian function $L(A,B,E,\lambda)$, given by

$$\begin{aligned} L =& \|A\|_* + \alpha\|E\|_1 + \beta\|B(I-Q)\|_F^2 + \gamma\|B^\top A\|_F^2 + \langle\lambda, X-A-B-E\rangle + \frac{\mu}{2}\|X-A-B-E\|_F^2 \\ =& \|A\|_* + \alpha\|E\|_1 + \beta\|B(I-Q)\|_F^2 + \gamma\|B^\top A\|_F^2 + \frac{\mu}{2}\|X-A-B-E+\frac{\lambda}{\mu}\|_F^2 - \frac{1}{2\mu}\|\lambda\|_F^2 \\ =& \|A\|_* + \alpha\|E\|_1 + \beta\|B(I-Q)\|_F^2 + h(A,B,E,\lambda,\mu) - \frac{1}{2\mu}\|\lambda\|_F^2\,, \end{aligned} \tag{5.5}$$

where $\langle X, Y\rangle \doteq \text{trace}(X^\top Y)$, μ is a positive scalar, λ is a Lagrange multiplier matrix, and $h(A,B,E,\lambda,\mu) = \frac{\mu}{2}\|X-A-B-E+\frac{\lambda}{\mu}\|_F^2 + \gamma\|B^\top A\|_F^2$ is a quadratic convenience function. We optimize (5.5) with an alternating direction strategy, and at every outer iteration of Algorithm 1, A, B, and E are first iteratively updated until convergence; subsequently, λ and μ are updated. The inner iteration updates of Algorithm 1 are given below.

Updating A_{k+1}: From the reduced augmented Lagrangian it is convenient to use the linearization technique of the LADMAP method [54], very effectively used also by other

approaches [36, 55, 18], and replace the quadratic term h with its first order approximation, computed at iteration k, and add a proximal term, giving the following update

$$\begin{aligned} A_{k+1} &= \arg\min_A \|A\|_* + < \nabla_A h(A_k, B_k, E_k, \lambda_k, \mu_k), A - A_k > + \frac{\eta\mu_k}{2}\|A - A_k\|_F^2 \\ &= \arg\min_A \|A\|_* + \frac{\eta\mu_k}{2}\|A - (X - B_k - E_k + \frac{\lambda_k}{\mu_k} - \gamma B_k B_k^\top A_k)\|_F^2 \ , \end{aligned} \tag{5.6}$$

where η must be greater than $\|A\|_F^2$ [54]. The solution to (5.6) is reported in Algorithm 1, and is obtained by applying the singular value thresholding algorithm [9], with the *soft-thresholding shrinkage operator* $\mathcal{S}_\epsilon(x)$, which is equal to: $x - \epsilon$ if $x > \epsilon$, $x + \epsilon$ if $x < -\epsilon$, and 0 elsewhere.

Updating E_{k+1}: From (5.5), the augmented Lagrangian reduces to

$$E_{k+1} = \arg\min_E \alpha\|E\|_1 + \frac{\mu_k}{2}\|E - (X - A_{k+1} - B_k + \frac{\lambda_k}{\mu_k})\|_F^2 \ , \tag{5.7}$$

Algorithm 1 Invariant Components Learning via the Exact ALM Method

Require: Observation matrix X, labels Q, and penalty weights α, β, γ

$k = 0;\ \rho > 1;\ \mu_0 > 0;\ \eta = \|X\|_F^2;\ \lambda_0 = \frac{\mathrm{sgn}(X)}{\max(\|\mathrm{sgn}(X)\|_F, \alpha^{-1}\|\mathrm{sgn}(X)\|_\infty)};\ A_0 = 0;\ B_0 = XQ;\ E_0 = 0$

while not converged **do**

  $j = 0;\ A_k^0 = A_k;\ B_k^0 = B_k;\ E_k^0 = E_k$

  while not converged **do**

  ▷ Line 5 solves (5.6)

    $(U, \Sigma, V) = \mathrm{svd}(X - B_k^j - E_k^j + \mu_k^{-1}\lambda_k - \gamma B_k^j {B_k^j}^\top A_k^j)\ ; \quad A_k^{j+1} = U\mathcal{S}_{(\eta\mu_k)^{-1}}(\Sigma)V^\top$

  ▷ Line 6 solves (5.7)

    $E_k^{j+1} = \mathcal{S}_{\alpha\mu_k^{-1}}(X - A_k^{j+1} - {B_k^j}^\top + \mu_k^{-1}\lambda_k)$

    Update B_k^{j+1} by solving (5.9) with A_k^{j+1} and E_k^{j+1}

    $j \leftarrow j + 1$

  end while

  $A_{k+1} = A_k^{j+1};\ B_{k+1} = B_k^{j+1};\ E_{k+1} = E_k^{j+1}$

  $\mu_{k+1} = \rho\mu_k; \quad \lambda_{k+1} = \lambda_k + \mu_k(X - A_{k+1} - B_{k+1} - E_{k+1})$

  $k \leftarrow k + 1$

end while

Ensure: A_k, B_k, E_k

and the solution, reported in Algorithm 1, is still obtained with an instance of the singular value thresholding algorithm [9].

Updating Y_{k+1}: This update is computed as

$$B_{k+1} = \arg\min_B \frac{\mu_k}{2}\|X - A_{k+1} - E_{k+1} - B + \frac{\lambda_k}{\mu_k}\|_F^2 + \beta\|B(I-Q)\|_F^2 + \gamma\|B^\top A_{k+1}\|_F^2 \ . \quad (5.8)$$

Note that the cost function in (5.8) is quadratic in B. Therefore, the update can be obtained by computing the partial derivative with respect to B of the cost function, and then setting it to zero. This leads to a Sylvester equation in B, given by

$$\gamma A_{k+1}A_{k+1}^\top B + B\left((\beta + \frac{\mu_k}{2})I - 2\beta Q - \beta QQ^\top\right) = \frac{\mu_k}{2}\left(D - A_{k+1} - E_{k+1} + \frac{\lambda_k}{\mu_k}\right) \ . \quad (5.9)$$

Therefore, the update (5.8) can be computed with a standard Sylvester equation solver. The full optimization procedure is summarized in Algorithm 1.

5.2 Classification

Given a test data point x, the obvious approach to perform classification is to compute a label via $y = \arg\min_i d(x, B_i)$, where $d(\cdot,\cdot)$ is a suitable distance between x and the invariant matrix B_i, representing class i. Following the strategy outlined in Section 4.2, from the invariant components B_i one can estimate $P_{\mathcal{B}_i} : \mathbb{R}^m \to \mathcal{B}_i$, the operator that projects data points directly onto $\mathcal{B}_i \subset \mathcal{B}$, the invariant subspace for class i. Doing so has the advantage that the projection of x onto $\mathcal{V}^\perp$ gives $b + P_{\mathcal{V}^\perp}e$, whereas the projection of x onto $\mathcal{B}_i$ gives $b + P_{\mathcal{B}_i}e$, and since $\mathcal{B}_i \subset \mathcal{V}^\perp$, it follows that $\|P_{\mathcal{B}_i}e\|_F \leq \|P_{\mathcal{V}^\perp}e\|_F$, which means a lower noise corruption. Therefore, we propose to use the following Frobenius norm $d_F(x, B_i) = n_i^{-1}\|B_i - P_{\mathcal{B}_i}x\,1_{n_i}^\top\|_F$. Note that if B_i can be approximated with $b_i 1_{n_i}^\top$, as it normally should, then the distance computation is even faster, because given by

$$d_F(x, B_i) = \|b_i - P_{\mathcal{B}_i}x\|_F \ . \quad (5.10)$$

We now make the following observation. Without loss of generality, let us assume that the columns of B are zero mean. The invariance constraint (5.3) can be re-written as

$Q = B^\top (BB^\top)^+ B$, where the covariance of B appears under the form of $BB^\top$ (for a short discussion there is no need to address the rank deficiency of B, and the use of the pseudoinverse $(BB^\top)^+$). Therefore, it is easy to realize that the Mahalanobis distance $d_M(b_i, b_j)$, between the invariant components b_i and b_j, for classes i and j is equal to 0 if $i = j$, and to $\sqrt{2N}$ if $i \neq j$, where for simplicity we have assumed $n_i = n_j$. This means that $\mathcal{B}$ is such that two different sufficient components s_i and s_j originate different (i.e., $b_i = s_i - P_{\mathcal{B}} s_i \neq s_j - P_{\mathcal{B}} s_j = b_j$), and equidistant (i.e., $d_M(b_i, b_j) = \sqrt{2N}\ \forall i \neq j$), invariant components, thus promoting a uniform class separation.

The observation above suggests the use of the Mahalanobis distance for testing, e.g., in the form of $d_M(x, B_i) = \sum_{b \in B_i} d_M(x, b)$. However, it is more convenient to use the corresponding similarity measure $\kappa(b_i, b_j) = b_i^\top (BB^\top)^+ b_j$, which gives 0 if $i \neq j$, and $\frac{1}{n_i}$ if $i = j$. Therefore, we propose the similarity measure defined as $\kappa(x, B_i) = 1_{n_i}^\top B_i^\top (BB^\top)^+ x$, and the label assignment is done according to $y = \arg\max_i \kappa(x, B_i)$. If $B_i = b_i 1_{n_i}^\top$, the similarity reduces to

$$\kappa(x, B_i) = n_i b_i^\top (BB^\top)^+ x \ . \tag{5.11}$$

Chapter 6

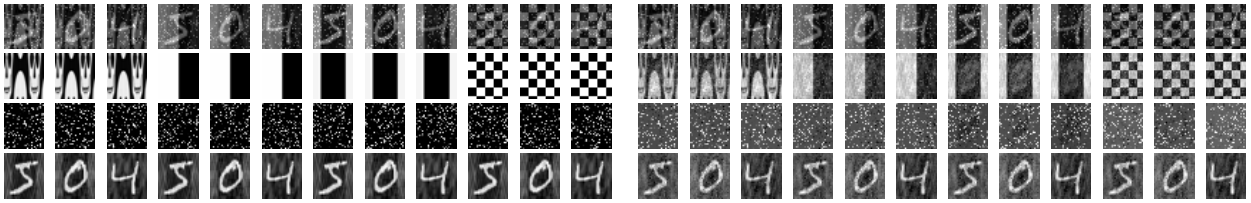
Experiments and Results

6.1 Experiments

In order to validate the proposed method we have performed experiments on synthetic data, on two face recognition datasets, and on one person reidentification dataset . All the results were obtained with a grid search of the parameters α, β, and γ.

6.1.1 Synthetic data.

To empirically verify the convergence of Algorithm 1, we have created a synthetic dataset made of $n = 120$ images of 32×28 pixels, with $N = 10$ invariant components depicting digits, and with image patterns representing A. The synthetic A and B satisfy the constraints (5.2), and (5.3), and we have added sparse noise E, corrupting 20% of randomly selected pixels, with values drawn from a uniform distribution between 0 and the largest possible pixel value in the image. Figure 6.1(a) shows the decomposition in A, E, and B of 12 synthetic data points, X (top row), and Figure 6.1(b) shows the estimated decomposition of the same points. Visually, the recovered decomposition closely resembles the originals, and the coefficients of variation (i.e., $\|\hat{z} - z\|_F / \|z\|_F$ where $\hat{z}$ is the estimated quantity), are 9.71%, 8.67%, 37.2%, for A, B, and E, respectively.



(a) (b)

Figure 6.1: **Synthetic data.** (a) Decomposition of 12 synthetic data points. (b) Decomposition of the same 12 points with Algorithm 1. Top row: input points X. Second row: A components. Third row: Sparse errors E. Bottom row: Invariant components B.

Approach	Recognition Rate
2DPCA [57]	96%
SRC [2]	93%
LRC [56]	93.5%
SLR [18]	93.5%
Our Method	95%

Table 6.1: Recognition Rate on AT&T face dataset.

6.1.2 AT&T Database

AT&T dataset is maintained at AT&T laboratories in Cambridge University, the dataset includes face images of 40 subjects taken in 10 controlled variations, which involves facial gestures (i.e. smiling and non-smiling) neutral or with glasses, also face rotational in four direction with no more that 20 degrees. similar to protocol in [56] we pick the first five images for training and last five images of each individual for testing. Table 6.1 illustrate the results for AT&T dataset. As it is illustrated in table its performance is in par with well known methods.

6.1.3 AR Dataset.

For this face recognition dataset [58], we follow a protocol used also by other recent works [1, 18]. The dataset contains over 4,000 frontal images of 126 people's faces (70 men and 56 women), images are taken in two sessions and under different facial expressions, illumination conditions and occlusions. In each session 3 images are occluded by sunglasses, 3 by a scarf,

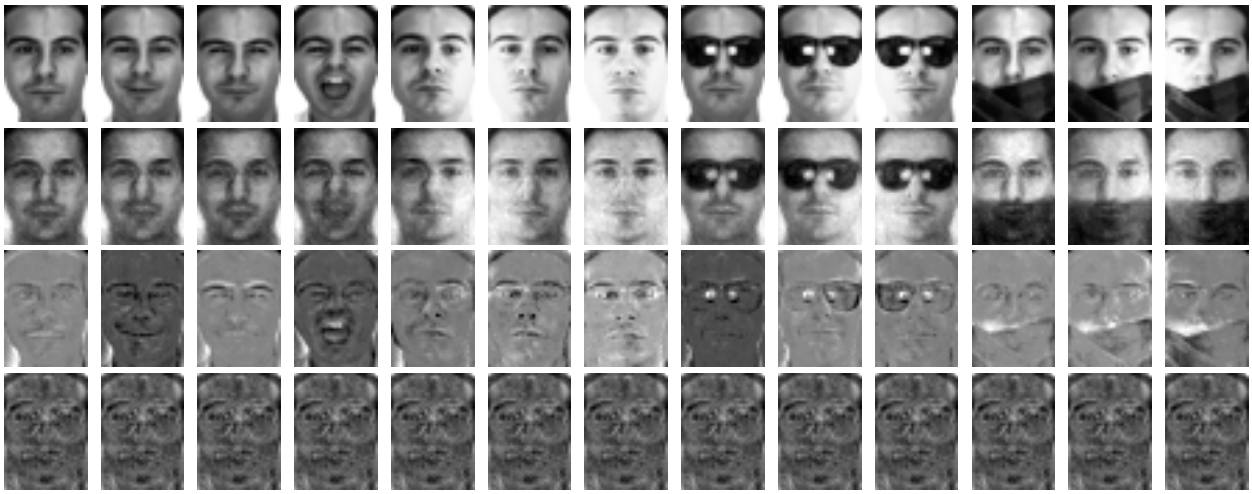


Figure 6.2: **AR dataset.** Decomposition results for the 13 images of one subject taken in one session. Row meanings are explained in Figure 6.1. Images are rescaled for better contrast and visualization.

dimention1230	p=1			p=2			p=3		
	sunglass	scarf	mixed	sunglass	scarf	mixed	sunglass	scarf	mixed
Our Method	**87.3**±0.30	**83.9**±0.52	**84.6**±0.37	**78.2**±0.32	**77.4**±0.35	**71.2**±0.52	**69.6**±0.62	**66.3**±0.53	**59.9**±0.47
SLR	87.1±0.64	83.0±0.57	81.8±0.70	76.1±0.78	73.8±0.79	66.3±0.98	60.5±0.98	58.2±1.07	51.2±1.26
LR w. Incoh.	86.8±0.40	82.9±0.34	79.1±0.56	73.9±0.49	72.3±0.52	65.4±0.75	58.4±0.72	58.1±0.88	49.9±0.89
SRC	84.9±0.23	76.2±0.42	79.2±0.42	73.2±0.44	70.8±0.30	63.6±0.64	56.2±0.80	60.1±0.75	46.2±1.06

Table 6.2: **Recognition Rate Comparison on AR** comparison between our methods and SLR [18], Low rank with incoherence [1] and SRC [2] on AR face dataset using the same protocol described in experiments under AR subsection

and are taken in different lighting conditions. The images are with $165 \times 120 = 19,800$ pixels, then converted into gray scale, and down-sampled by 4×4. As other authors did [1, 18, 53], we select a subset of 50 men and 50 women. Figure 6.2 illustrates 13 images taken from one subject in one session, along with the decomposition. The proposed algorithm effectively extracts the invariant component (bottom row), wich is pretty much identical for every image. The second row from top is a low-rank representation of the face images, and the second row from bottom is sparse noise.

Following [18, 1] we consider three scenarios, indicated as SUNGLASSES, SCARF and SUNGASSES+SCARF, where we do face recognition with highly corrupted training and testing data. For SUNGLASSES a subject in the training set is composed by p randomly selected face images occluded with sunglasses, and $8-p$ neutral, all selected from session 1. The remaining $6-p$ images occluded by sunglasses plus $6+p$ neutral from both sessions, form 12 testing

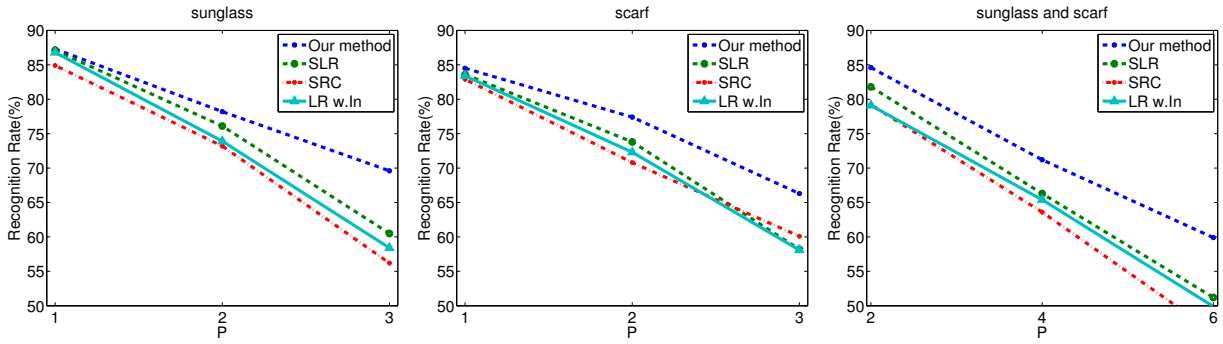


Figure 6.3: **AR dataset.** Recognition rates versus different numbers p, of corrupted training images per class for the three scenarios: sunglasses (left), scarf (center), sunglasses and scarf (right).

images per person. Note that face images with sunglasses are occluded about 20%. For the SCARF scenario, the data subdivision is identical only that we consider the face images occluded by a scarf, which produces occlusions of about 40%. For the SUNGLASSES+SCARF case, the difference is that for a given person, p images are occluded with sunglasses and p with the scarf, leaving 17 images for testing per person. Unlike previous work, that have shown results only for $p = 1$, here we also test the case for $p = 2$ and $p = 3$. The experiment has been repeated 5 times and the average recognition rates are plotted in Figure 6.3. The optimal penalty parameters were $\alpha = 1.5$, $\beta = 1000$, $\gamma = 0.9$. Unless otherwise specified, every result obtained in this section is with the distance (5.10). Along with ours, we have also tested the structured low-rank representation (SLR) approach [18], the low-rank with incoherence (LRwIn) approach [1], and the SRC [2]. We have reimplemented the SLR and the LRwIn approaches. For the SRC we have used the code publicly available. Every approach was tested with input images with the same size, and with other parameters set at the peak of their performance. From Figure 6.3 it can be appreciated that the proposed approach demonstrates a superior robustness with respect to corruption in the training set as p increases. For instance, compared to the overall best competitor, which is SLR, for the SUNGLASSES+SCARF case, for $p = 1$ the improvement is 2.8%, for $p = 2$ is 4.9%, and for $p = 3$ is 8.7%.

6.1.4 Extended-Yale B Dataset.

This face recognition dataset [59] contains tightly cropped face images of 38 subjects. Each of them has around 59 to 64 images taken under varying lighting conditions, which in total add up to 2,414 images. The cropped images are with $192 \times 168 = 32,256$ pixels. We randomly select 8, and in a subsequent experiment 32, training images for each person, and use the rest for testing in a recognition experiment. We repeat this 5 times and report the average recognition rate for the images down-sampled by a factor of 2, 4, and 8. For each of those conditions we also compare against the SLR [18], the LRwIn [1], and the SRC [2] approaches at the peak of their performance. For our approach the optimal penalty parameters were $\alpha = 0.9$, $\beta = 1000$, $\gamma = 0.01$. Figure 6.5 illustrates the comparison between the recognition rates. For the SRC, we also include what happens when the training set drops in size from 8 to 5, and from 32 to 20 training images. This experiment highlights that our approach compares favorably with the others especially when a smaller corrupted training dataset is available, and works on par with others (SRL and LRwIn) with lots of training data. This is because our approach inherently attempts learning a global variation space, shared by all the training data. So, even with fewer training images per person their aggregation allows learning the variation space better than in other approaches.

Figure 6.5, right, also shows a comparison between the two distances (5.10) and (5.11) on a subset of the dataset, with 32 training data points per person, against different image resolutions. Although in our experiments sometimes we found (5.11) to work better, the

Figure 6.4: **Extended Yale dataset.** Decomposition results for 2 subjects under 8 different illumination conditions. Row meanings are explained in Figure 6.1. Images are rescaled for better contrast and visualization.

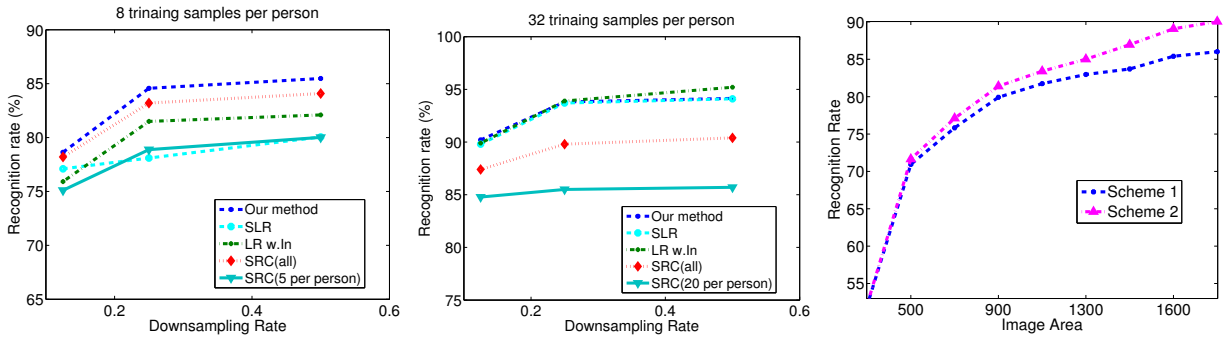


Figure 6.5: **Extended Yale B dataset.** From left to right: Recognition scores at different image downsampling rates for 8 and 32 training samples per subject; recognition rates obtained with the distance (5.10) (Schema 1) and the distance (5.11) (Schema 2) at various image resolutions and 32 training samples;

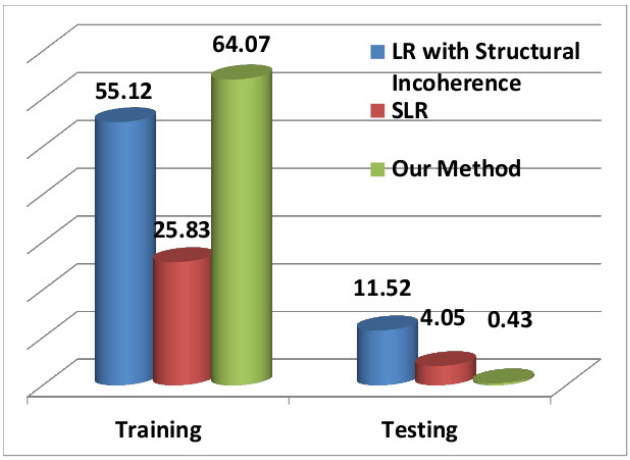


Figure 6.6: **Extended Yale B dataset.** Running time in seconds of our Matlab implementations for training and testing.

Frobenius option (5.10), like in this case, appeared more robust.

Running time when doing testing is linear in number of subjects in the training data and every distance we need to compute is just a dot product. Figure 6.6, shows a running time comparison between the Matlab implementations of ours, the SLR, and the LRwIn methods, running on a high-end PC. Our training procedure appears slightly more costly than the others, but, as anticipated, testing appears faster than SLR by a factor of 10, and faster than LRwIn by a factor of 25.

6.1.5 i-LIDS MCTS Dataset.

This dataset [49] contains 476 whole body person images of 119 people captured by multiple non overlapping surveillance cameras. There are 4 images on average per person. We excluded subjects who had only 1 or 2 images. All the images are normalized to 128×64 pixels. This dataset is used for person reidentification across camera views. Unlike faces that can be aligned, people images from unconstrained environments are highly misaligned, and this dataset pushes the proposed approach beyond limits. Nevertheless, Figure 6.7 shows on left the decomposition of the 3 training images of two people, and on the right the cumulative matching curves (CMC) with 30 and with 80 people in the training set. In a CMC curve, a rank r matching rate indicates the percentage of test (or probe) images with correct matches found in the top r ranks against the people in the training (or gallery) set. The penalty parameters were $\alpha = 1$, $\beta = 100$, $\gamma = 1$ for 30 subjects, and $\alpha = 2$, $\beta = 100$, $\gamma = 0.1$ for 80 subjects. To compare our results with two state-of-the-art approaches, namely the relative distance comparison (RDC) [49], and SDALF [43], we have used 3 images in the training set, and 1 image in the testing set per person. We run the experiment four times to make sure an image is on average part of the probe set once. For SDALF we learned the signature from 3 images in the training set for fair comparison. Despite the extreme conditions, to our surprise, from the CMC curves the proposed model is capable of keeping up with state-of-the-art approaches for higher matching ranks.

6.1.6 CAVIAR4REID

CAVIAR4REID is a new dataset for evaluating person re-identification algorithms. As the name suggests, the dataset has been extracted from the CAVIAR dataset mostly famous for person tracking and detection evaluations. It is a challenging dataset because it has broad changes in resolution and it is extracted from a real scenario where re-identification is necessary due to the presence of multiple cameras and the pose variations between the images are severe. For this dataset we carried our experiments in a similar fashion like i-LIDS dataset.

Figure 6.7: **i-LIDS MCTS dataset.** Decomposition results for 2 subjects under 4 different viewpoints. Row meanings are explained in Figure 6.1. Images are rescaled for better contrast and visualization.

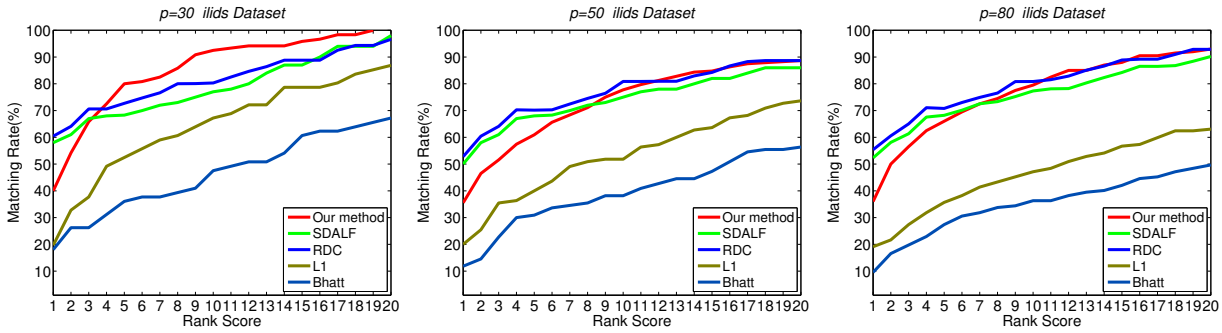


Figure 6.8: **CMC Curve** a) p = 30 b) p= 50 c) p =80 Performance comparison using CMC curves on the i-LIDS MCTS dataset

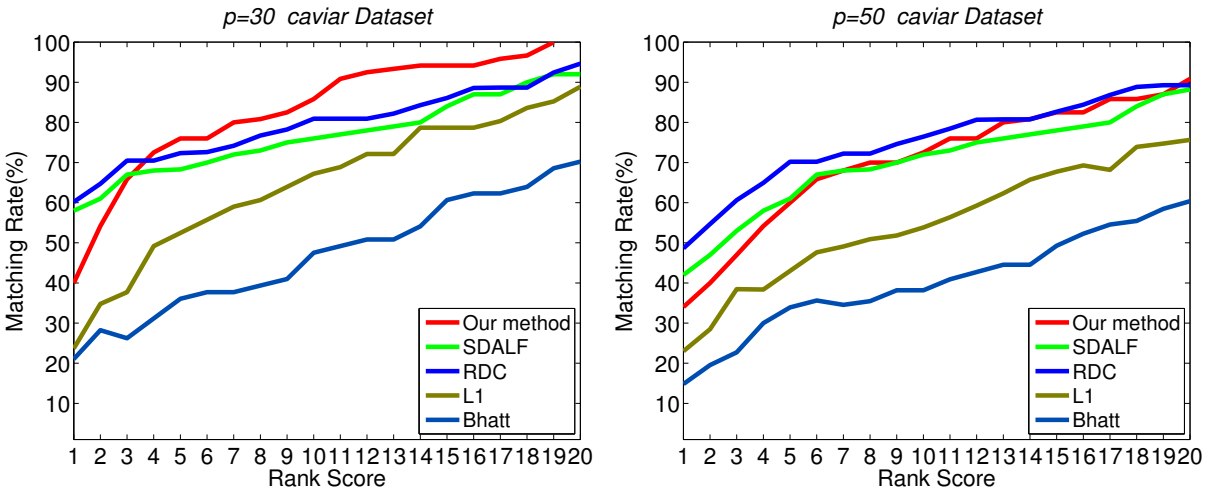


Figure 6.9: **CMC Curve** a) p = 30 b) p= 50 Performance comparison using CMC curves on the CAVIAR4REID dataset

Chapter 7

Conclusion

In this work we have introduced a invariance subspace representation. He vae used this novel representation to address the problem of human identification which contains grossly corrupted training and testing data. We have formulated our problem by extending the Robust PCA problem and included label information into the problem to make the learning a supervised one. Using the techniques of ALM we were able to solve this problem. We have used our approach to do face recognition and person re-identification. In case of face recognition we were able to out perform the state-of-the art approaches. Although in case of person re-identification we were competitive but this is due to the fact the approach is not designed in a fashion that can cope with huge alignment changes as present in this problem. One of the main advantages of our approach is that we can use simple geometry to do recognition. This allows us to improve the running time for doing testing.

We can take this work further ahead. One of the things than can be experimented is to adapt and extend the problem to do image classification instead of just doing human identification. Another thing that can be taken care of is that although our approach is efficient at test time we can try to improve the iterative algorithm to improve the training time computational complexity.

References

[1] Chih-Fan Chen, Chia-Po Wei, and Y.-C.F. Wang, "Low-rank matrix recovery with structural incoherence for robust face recognition," in *Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on*, June 2012, pp. 2618–2625.

[2] J. Wright, AY. Yang, A Ganesh, S.S. Sastry, and Yi Ma, "Robust face recognition via sparse representation," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 31, no. 2, pp. 210–227, Feb 2009.

[3] Meng Yang and Lei Zhang, "Gabor feature based sparse representation for face recognition with gabor occlusion dictionary," in *Computer Vision ECCV 2010*, Kostas Daniilidis, Petros Maragos, and Nikos Paragios, Eds., vol. 6316 of *Lecture Notes in Computer Science*, pp. 448–461. Springer Berlin Heidelberg, 2010.

[4] Michael Grant and Stephen Boyd, "CVX: Matlab software for disciplined convex programming, version 2.1," `http://cvxr.com/cvx`, Mar. 2014.

[5] Venkat Chandrasekaran, Sujay Sanghavi, Pablo A. Parrilo, and Alan S. Willsky, "Rank-sparsity incoherence for matrix decomposition," *SIAM Journal on Optimization*, vol. 21, no. 2, pp. 572–596, 2011.

[6] Elaine T. Hale, Wotao Yin, and Yin Zhang, "Fixed-point continuation for l1-minimization: Methodology and convergence," .

[7] Wotao Yin, Stanley Osher, Donald Goldfarb, and Jerome Darbon, "Bregman iterative algorithms for l1-minimization with applications to compressed sensing," *SIAM J. Imaging Sci*, pp. 143–168, 2008.

[8] Amir Beck and Marc Teboulle, "A fast iterative shrinkage-thresholding algorithm for linear inverse problems," *SIAM J. Img. Sci.*, vol. 2, no. 1, pp. 183–202, Mar. 2009.

[9] Jian-Feng Cai, Emmanuel J. Candès, and Zuowei Shen, "A singular value thresholding algorithm for matrix completion," *SIAM J. on Optimization*, vol. 20, no. 4, pp. 1956–1982, Mar. 2010.

[10] Emmanuel Candès and Benjamin Recht, "Exact matrix completion via convex optimization," *Commun. ACM*, vol. 55, no. 6, pp. 111–119, June 2012.

[11] Benjamin Recht, Maryam Fazel, and Pablo A. Parrilo, "Guaranteed minimum-rank solutions of linear matrix equations via nuclear norm minimization," *SIAM Rev.*, vol. 52, no. 3, pp. 471–501, Aug. 2010.

[12] John Wright, Arvind Ganesh, Shankar Rao, Yigang Peng, and Yi Ma, "Robust principal component analysis: Exact recovery of corrupted low-rank matrices via convex optimization," in *Advances in Neural Information Processing Systems 22*, Y. Bengio, D. Schuurmans, J.D. Lafferty, C.K.I. Williams, and A. Culotta, Eds., pp. 2080–2088. Curran Associates, Inc., 2009.

[13] Zhouchen Lin, Arvind Ganesh, John Wright, Leqin Wu, Minming Chen, and Yi Ma, "Fast convex optimization algorithms for exact recovery of a corrupted low-rank matrix," in *In Intl. Workshop on Comp. Adv. in Multi-Sensor Adapt. Processing, Aruba, Dutch Antilles*, 2009.

[14] Z. Lin, M. Chen, and Y. Ma, "The Augmented Lagrange Multiplier Method for Exact Recovery of Corrupted Low-Rank Matrices," *ArXiv e-prints*, Sept. 2010.

[15] Bin Cheng, Guangcan Liu, Jingdong Wang, Zhongyang Huang, and Shuicheng Yan, "Multi-task low-rank affinity pursuit for image segmentation," in *Computer Vision (ICCV), 2011 IEEE International Conference on*, Nov 2011, pp. 2439–2446.

[16] Ying Wu, "A unified approach to salient object detection via low rank matrix recovery," in *Proceedings of the 2012 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Washington, DC, USA, 2012, CVPR '12, pp. 853–860, IEEE Computer Society.

[17] Chunjie Zhang, Jing Liu, Qi Tian, Changsheng Xu, Hanqing Lu, and Songde Ma, "Image classification by non-negative sparse coding, low-rank and sparse decomposition," in *Computer Vision and Pattern Recognition (CVPR), 2011 IEEE Conference on*, June 2011, pp. 1673–1680.

[18] Yangmuzi Zhang, Zhuolin Jiang, and Larry S. Davis, "Learning structured low-rank representations for image classification.," in *CVPR*, 2013, pp. 676–683.

[19] Zhengdong Zhang, Y. Matsushita, and Yi Ma, "Camera calibration with lens distortion from low-rank textures," in *Computer Vision and Pattern Recognition (CVPR), 2011 IEEE Conference on*, June 2011, pp. 2321–2328.

[20] Zhengdong Zhang, Arvind Ganesh, Xiao Liang, and Yi Ma, "Tilt: Transform invariant low-rank textures," *International Journal of Computer Vision*, vol. 99, no. 1, pp. 1–24, 2012.

[21] Zhengdong Zhang, Xiao Liang, and Yi Ma, "Unwrapping low-rank textures on generalized cylindrical surfaces," in *Computer Vision (ICCV), 2011 IEEE International Conference on*, Nov 2011, pp. 1347–1354.

[22] Lun Wu, Arvind Ganesh, Boxin Shi, Yasuyuki Matsushita, Yongtian Wang, and Yi Ma, "Robust photometric stereo via low-rank matrix completion and recovery," in *Proceedings of the 10th Asian Conference on Computer Vision - Volume Part III*, Berlin, Heidelberg, 2011, ACCV'10, pp. 703–717, Springer-Verlag.

[23] Yigang Peng, A. Ganesh, J. Wright, Wenli Xu, and Yi Ma, "Rasl: Robust alignment by sparse and low-rank decomposition for linearly correlated images," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 34, no. 11, pp. 2233–2246, 2012.

[24] Guangcan Liu, Zhouchen Lin, Shuicheng Yan, Ju Sun, Yong Yu, and Yi Ma, "Robust recovery of subspace structures by low-rank representation," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 35, no. 1, pp. 171–184, Jan 2013.

[25] Risheng Liu, Zhouchen Lin, Fernando De la Torre, and Zhixun Su, "Fixed-rank representation for unsupervised visual learning," in *CVPR*, 2012, pp. 598–605.

[26] Ehsan Elhamifar and René Vidal, "Sparse subspace clustering: Algorithm, theory, and applications," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 11, pp. 2765–2781, 2013.

[27] V.M. Patel, Hien Van Nguyen, and R. Vidal, "Latent space sparse subspace clustering," in *Computer Vision (ICCV), 2013 IEEE International Conference on*, Dec 2013, pp. 225–232.

[28] M.A Turk and AP. Pentland, "Face recognition using eigenfaces," in *Computer Vision and Pattern Recognition, 1991. Proceedings CVPR '91., IEEE Computer Society Conference on*, Jun 1991, pp. 586–591.

[29] Peter N. Belhumeur, Joo P. Hespanha, and David J. Kriegman, "Eigenfaces vs. fisherfaces: Recognition using class specific linear projection," 1997.

[30] Xiaofei He, Shuicheng Yan, Yuxiao Hu, P. Niyogi, and Hong-Jiang Zhang, "Face recognition using laplacianfaces," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 27, no. 3, pp. 328–340, March 2005.

[31] Fernando De La Torre and Michael J. Black, "A framework for robust subspace learning," *Int. J. Comput. Vision*, vol. 54, no. 1-3, pp. 117–142, Aug. 2003.

[32] Qifa Ke and Takeo Kanade, "Robust l1 norm factorization in the presence of outliers and missing data by alternative convex programming," in *IEEE CONF. COMPUTER VISION AND PATTERN RECOGNITION*, 2005, pp. 592–599.

[33] Emmanuel J. Candès, Xiaodong Li, Yi Ma, and John Wright, "Robust principal component analysis?," *J. ACM*, vol. 58, no. 3, pp. 11:1–11:37, June 2011.

[34] A. Wagner, J. Wright, A. Ganesh, Zihan Zhou, and Yi Ma, "Towards a practical face recognition system: Robust registration and illumination by sparse representation," in *Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on*, June 2009, pp. 597–604.

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

[36] Long Ma, Chunheng Wang, Baihua Xiao, and Wen Zhou, "Sparse representation for face recognition based on discriminative low-rank dictionary learning," in *Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on*, June 2012, pp. 2586–2593.

[37] U. Park, AK. Jain, I Kitahara, K. Kogure, and N. Hagita, "Vise: Visual search engine using multiple networked cameras," in *Pattern Recognition, 2006. ICPR 2006. 18th International Conference on*, 2006, vol. 3, pp. 1204–1207.

[38] Weiming Hu, Min Hu, Xue Zhou, Tieniu Tan, Jianguang Lou, and Steve Maybank, "Principal axis-based correspondence between multiple cameras for people tracking," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 28, no. 4, pp. 663–671, Apr. 2006.

[39] Douglas Gray and Hai Tao, "Viewpoint invariant pedestrian recognition with an ensemble of localized features," in *Proceedings of the 10th European Conference on Computer Vision: Part I*, Berlin, Heidelberg, 2008, ECCV '08, pp. 262–275, Springer-Verlag.

[40] Xiaogang Wang, G. Doretto, T. Sebastian, J. Rittscher, and P. Tu, "Shape and appearance context modeling," in *Computer Vision, 2007. ICCV 2007. IEEE 11th International Conference on*, Oct 2007, pp. 1–8.

[41] N. Gheissari, T.B. Sebastian, and R. Hartley, "Person reidentification using spatiotemporal appearance," in *Computer Vision and Pattern Recognition, 2006 IEEE Computer Society Conference on*, 2006, vol. 2, pp. 1528–1535.

[42] P. Dollar, Zhuowen Tu, Hai Tao, and S. Belongie, "Feature mining for image classification," in *Computer Vision and Pattern Recognition, 2007. CVPR '07. IEEE Conference on*, June 2007, pp. 1–8.

[43] M. Farenzena, L. Bazzani, A. Perina, V. Murino, and M. Cristani, "Person reidentification by symmetry-driven accumulation of local features," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2010, pp. 2360 –2367.

[44] Rui Zhao, Wanli Ouyang, and Xiaogang Wang, "Unsupervised salience learning for person re-identification," in *Computer Vision and Pattern Recognition (CVPR), 2013 IEEE Conference on*, June 2013, pp. 3586–3593.

[45] Ryan Layne, Timothy M. Hospedales, and Shaogang Gong, "Towards person identification and re-identification with attributes," in *Proceedings of the 12th International Conference on Computer Vision - Volume Part I*, Berlin, Heidelberg, 2012, ECCV'12, pp. 402–412, Springer-Verlag.

[46] Chunxiao Liu, C.C. Loy, Shaogang Gong, and Guijin Wang, "Pop: Person reidentification post-rank optimisation," in *Computer Vision (ICCV), 2013 IEEE International Conference on*, Dec 2013, pp. 441–448.

[47] Kilian Q. Weinberger and Lawrence K. Saul, "Distance metric learning for large margin nearest neighbor classification," *J. Mach. Learn. Res.*, vol. 10, pp. 207–244, June 2009.

[48] Amir Globerson and Sam Roweis, "Metric learning by collapsing classes," .

[49] Wei-Shi Zheng, Shaogang Gong, and Tao Xiang, "Person re-identification by probabilistic relative distance comparison," in *Proceedings of the 2011 IEEE Conference on Computer Vision and Pattern Recognition*, Washington, DC, USA, 2011, CVPR '11, pp. 649–656, IEEE Computer Society.

[50] A Mignon and F. Jurie, "Pcca: A new approach for distance learning from sparse pairwise constraints," in *Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on*, June 2012, pp. 2666–2672.

[51] H. Hotelling, "Analysis of a complex of statistical variables into principal components," *J. Educ. Psych.*, vol. 24, 1933.

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

[53] Weihong Deng, Jiani Hu, and Jun Guo, "In defense of sparsity based face recognition," in *Computer Vision and Pattern Recognition (CVPR), 2013 IEEE Conference on*, June 2013, pp. 399–406.

[54] Zhouchen Lin, Risheng Liu, and Zhixun Su, "Linearized alternating direction method with adaptive penalty for low-rank representation," in *NIPS*, 2011, pp. 612–620.

[55] Liansheng Zhuang, Haoyuan Gao, Zhouchen Lin, Yi Ma, Xin Zhang, and Nenghai Yu, "Non-negative low rank and sparse graph for semi-supervised learning," in *Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on*, June 2012, pp. 2328–2335.

[56] I. Naseem, R. Togneri, and M. Bennamoun, "Linear regression for face recognition," *IEEE TPAMI*, vol. 32, no. 11, pp. 2106–2112, Nov 2010.

[57] J. Yang, D. Zhang, A. F. Frangi, and J.Y. Yang, "Two-dimensional PCA: a new approach to appearance-based face representation and recognition," *IEEE TPAMI*, vol. 26, no. 1, pp. 131–137, 2004.

[58] Aleix M Martinez, "The ar face database," *CVC Technical Report*, vol. 24, 1998.

[59] Athinodoros S. Georghiades, Peter N. Belhumeur, and David J. Kriegman, "From few to many: Illumination cone models for face recognition under variable lighting and pose," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 23, no. 6, pp. 643–660, June 2001.