FAST AND ROBUST FACE RECOGNITION

by

Aliya Zafar

2010-NUST-MS PHD- MTS-06

MS-66 (MTS)

Submitted to the Department of Mechatronics Engineering in fulfillment of the requirements for the degree of

MASTER OF SCIENCE

in

MECHATRONICS ENGINEERING

Thesis Supervisor

Dr Javaid Iqbal

College of Electrical & Mechanical Engineering

National University of Sciences & Technology

2013

In the name of Allah, the most

Beneficent and the most Merciful

DECLARATION

I hereby declare that I have developed this thesis entirely on the basis of my personal efforts under the sincere guidance of my co-supervisor (Dr. Rab Nawaz). All the sources used in this thesis have been cited and the contents of this thesis have not been plagiarized. No portion of the work presented in this thesis has been submitted in support of any application for any other degree of qualification to this or any other university or institute of learning.

Aliya Zafar

ACKNOWLEDGEMENTS

Thanks to Allah (s.w.t.) for enabling me to achieve this milestone. I am grateful to my husband, his aunt and both our parents for their encouragement, immense love, prayers and moral support throughout my academic career.

I am deeply beholden to my co-supervisor, Dr. Rab Nawaz, for his continuous guidance, patiently delivering his ideas and giving time for the fruitful and long sessions. His way of teaching complicated things with simplest examples and excellent programming skills will be helpful for me in my practical life. I am also grateful to Head of Department (Dr. Javaid Iqbal), for his continuous support during my thesis work. The valuable suggestions and comments of all the teachers present during pre-defense were helpful to improve the research work presented in this thesis.

ABSTRACT

Face-Recognition is one of the most efficient and highly popular technology in biometrics because of its least intrusiveness, reliability and easy use. We introduce a novel algorithm namely, **Robust NCC**, for face recognition with varying illumination, expression, occlusion, disguise.

RNCC has exceptionally remarkable results. It’s fast and can correctly identify highly occluded images without adding much complexity.

To deal with illumination variation, the accuracy rate is further improved by cascading it with Collaborative Representation Classifier (CRC). In order to use the two classifiers in fusion setting, we perform intelligent cascading through confidence weighted scheme.

The final cascaded method is tested on 7 renowned databases (AR, Extended Yale B, Cohn Kanade and Cohn Kanade plus, Bosphorus, Yale Faces, Jaffe). It outperforms state of the art Sparse Representation and other well-known classifiers. The recognition rate for all the tested databases with low dimensionality (13 x 10) is above 90%.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

$\mathbb{R}$	Linear subspace
$\mathbb{R}^n$	n dimensional image space
$\mathbb{R}^{m \times n}$	Image space with m rows and n columns
SRC	Sparse Representation Classifier
CRC_RLS	Collaborative Representation Classifier with Regularized Least Square
RNCC	Robust Normalized Cross Correlation
FRS	Face Recognition System

Chapter 1

Introduction

1.1 Biometrics

Biometrics uses unique identifiers such as voice inflections in speech, fingerprints, various iris patterns, hand shape/geometry and blood vessel patterns in the retina to ascertain and verify people's identity.

1. Major Biometric Technologies

Iris, hand geometry and fingerprint recognition are the most commonly used technologies. Other non-common ones are facial, keystroke, signature and voice recognition. However, important biometric technologies of the future include face, DNA, earlobe and gait recognition.

2. Biometric Advantages

- Can't be forged
- Positive & accurate Identification
- No more stolen or forgotten passwords
- Serves as a user friendly & non-transferable Key
- Highest level of security
- Offers mobility
- Identify the individual in spite of time variation (it does not matter if the first biometric sample was taken year ago)

3. Common Human Biometric Characteristics

Biometric technologies are further subdivided in two categories, namely:

- behavioral patterns (keyboard typing, gait, hand-grip, signature, lip movement)
- physiological traits (body odor, face, voice, body salinity, iris, vascular, fingerprint, ear shape, hand geometry)

4. Soft & Hard Biometrics

Traits which are not distinctive enough to identify an individual uniquely are known as Soft biometrics [1] e.g.

- physical: skin color, eye color, hair color, presence of beard, presence of moustache, height, weight
- behavioral: gait, keystroke
- Adhered human characteristics: clothes color, tattoos, accessories

Traits which are unique to an individual are called Hard biometrics.

5. Verification And Identification

Face recognition problem has two main stages:

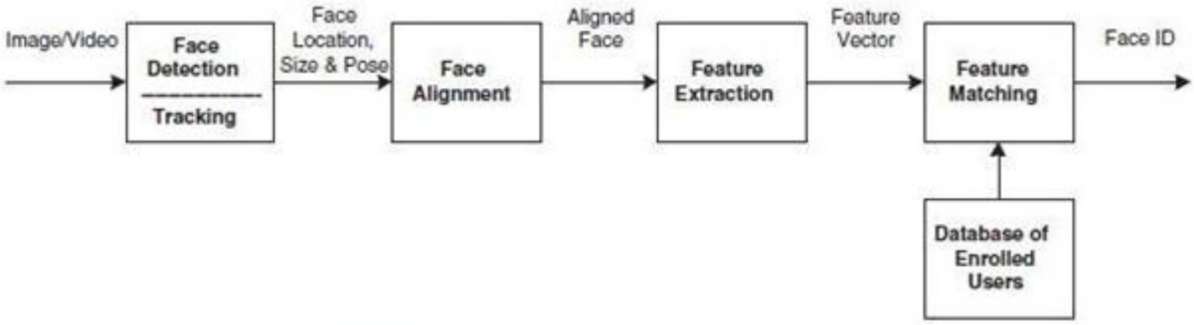

Figure 1: Face recognition processing flow

- An identification/recognition system performs 1:N matching i.e. it compares a test subject with ‘N’ other subjects
- In contrast, an authentication/verification system performs 1:1 matching i.e. it confirms or denies the identity claimed by a test subject

Both the above mentioned techniques have the same initial steps and a major portion of the classifier design. However their applications are distinct.

1.2 Biometric Face Recognition

Face recognition is the technique for figuring out whose face it is.

1. Advantages over other biometrics

Face recognition is considered as one of the most efficient technology as compared to other biometric technologies. For example, Iris recognition devices expose the diseases which one may not wish to reveal. Fingerprint devices can be hacked and duplicated. Also, access can be denied in case of dirty or wet fingers. Face recognition system is better than the rest in various aspects. It is a friendly, non-intrusive and passive system to verify identity of a person.

2. *Applications*

The largest face recognition systems in the world, operates in the U.S. Department of State that is actively used for visa processing with over 75 million photographs. Most recently, the US Federal Bureau of Investigation (FBI) has launched $1 billion nationwide facial recognition system along with voice samples, DNA records, iris scans, and other biometrics, expected to be rolled out nationwide by 2014. This will greatly help the FBI for catching criminals.

FRS technology has reached the point where it can correctly can match a single face among 1.6 million passport or mugshots images in less than 1.2 seconds with 92% accuracy [2]. It is an expedient way to correctly identify suspects where iris or DNA records exist. Images captures in low light are the most difficult ones to match.

Most recently, face recognition is being used in various softwares and websites e.g. Facebook, Google Picasa, Sony's Picture Motion Browser (PMB), Windows Live Photo Gallery, etc.

3. Weaknesses vs. Strengths

Facial recognition may not be the most efficient and reliable among the different biometric techniques but it has several advantages over the rest. Other biometrics like iris, speech recognition and fingerprints cannot perform such non-intrusive mass scanning. Through FRS installed in multiplexes, airports, and other public places presence of criminals among the crowd can be detected. However, its effectiveness has been severely questioned for security usage.

Despite the successes of many systems, many issues remain to be addressed. Among those issues, the following are prominent for most systems: expression, illumination and pose problem, images captured years apart, scale variation, partially occluded faces (by moustaches, glasses, beards), low quality image acquisition, twins etc. Another significant problem other than recognizing images is how to compare different face recognition systems.

(a)

(b)

(c)

Figure 2: Variation for the same face: (a) Illumination (b) Expression (c) Pose

The illumination, expression and pose problems are represented in the Fig. 2, where the changes induced due to illumination, expression and pose might be greater than the differences between subjects, which results in misclassifying the test subject.

Many fear that face recognition threatens anonymity in public. Inherent dangers are associated with collection of data. Using the right software, a person's name and address can be found in few seconds no matter where his face is photographed. Social networking platforms capture the biometric data of a person when he is tagged (identified in an online image) and the person's name and details can be provided when the image shows up online again.

According to a presentation at Black Hat security conference [23], online anonymity is dead. The worst thing is there are files that store the data of millions of people. It is possible for both private enterprise and state to use and abuse this data by identifying people through photographs.

1.3 2D vs. 3D Face Recognition

Both 2D and 3D FRSs have their cons and pros. A comparison of 2D FRS vs. 3D FRS is given below:

2D images are extremely common due to easy availability of 2D digital cameras (e.g. in mobile phones).	3D FRS is a lot more expensive than 2D FRS; therefore, it's rarely used.
A 2D face image or database accounts only two dimensions of the face.	As face is a 3D object, therefore, a much better performance, beating 2D accuracy rate with a large margin is expected from it. However, this popular belief has not been proved by any experiment till now.
2D is not robust to light variations e.g. a single face may have various models created by different sources of light.	3D data capturing has similar traits.
2D represents a face by intensity variation. It discriminates faces by comparing intensity or color of the features of a given face.	3D represents a face by shape variation. It discriminates faces on the basis of the shape of the features of a given face.
Constant improvements in the 2D FRS.	3D devices in the market are not as matured as 2D devices.
2D FRS cannot handle pose of the head problem.	3D FRS can tackle pose of the head. Also the surface curvature of the head can now be used to describe a face.

Another advantage of this category is that it adds depth information to conventional statistical appearance-based methods (like PCA, LDA, ICA…) without increasing too much of the computational cost. According to a survey [3], best method is to use multi-modal recognition (2D+3D) as it allows greater accuracy than either modality alone.

1.4 Problem Statement

In this thesis, we aim at developing a novel 2D facial recognition system that is robust to illumination, expression, occlusion and disguise. There are four major parts of this research work:

- propose a novel, fast and robust facial recognition system
- improve previously stated results in 2D Face Recognition
- test on renowned databases e.g. AR, Extended Yale B, Bosphorous, Yale Faces, Jaffe and Extended Cohn Kanade Database
- compare results with the other state of the art techniques

Chapter 2

Literature Review

2.1 Introduction

FR has been a strong field of research since 1990s and more techniques are being invented each year but it is still far from reliable.

1. How Recognition Works

Suppose 'N' face images and a test face image is given to be recognized:

- Find "distance" between test and each of the 'N' face images
- Choose the image with minimum distance as the best match
- Further compare it with a threshold, if distance > threshold, choose the image with minimum distance as the best match, otherwise classify as unknown test image

2. How "Far Apart" Are Test and Training Images

Distance between two points say P_1 and P_2 is called Euclidean distance [4]. In 2D,

$$d_{12} = sqrt(dx^2 + dy^2) \tag{1}$$

where $dx = x_2 - x_1$ and $dy = y_2 - y_1$. In 3D, it's $sqrt(dx^2 + dy^2 + dz^2)$.

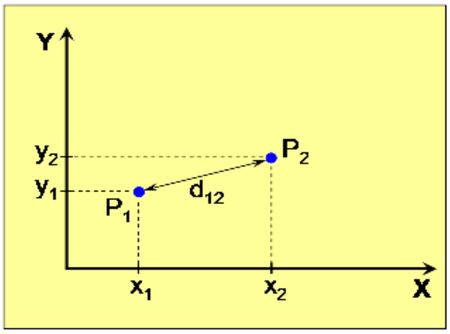

Figure 3: In 2D, Euclidean distance (d_{12}) for two points

*3. Noise Times 1,600 Is A **Lot** Of Noise*

No two images are absolutely same e.g. a pixel may darken or brighten by just a minor (incidental) influence which acts as noise. If each one of the 1,600 pixels obtained by computing distance between 40x40 face images causes minor noise, the overall noise component will be very hugh.

According to the above assumption, summation of all the squared pixel differences would result in a hugh noise contribution that's extremely high in comparison to the fruitful/useful information that is actually required. The purpose of downsampling/dimensionality reduction is to reduce the noise level, so the useful information can be extracted.

The interesting/useful information usually lies in a much lower dimension than the actual one. Each pixel's value is a representative of a measurement in case of an image. Mostly, the information required to distinguish between faces from different individuals can be (somehow) represented with much value than 1,600. The actual required parameters number might be near 500 or just 50.

4. Criteria's To Evaluate The Effectiveness Of An Algorithm

According to a face recognition survey of 2007, the criteria's to evaluate the effectiveness of an algorithm are:

- For thorough assessment, use greater number of databases
- Use smaller dimension of the input images in order to speed up the system and reduce computational burden
- Recognition is still possible on 18 x 25 grey scale images (Kurita et al., 2003). Techniques using minimum dimension of input images are better as low resolution images are provided by most cameras used for video surveillance applications
- Identification is performed many times but in real situations, just one image is available as training. Therefore, the ratio gallery size/probe size should be minimized

5. Types Of Classification

Recognition process has been developed into a science of complex mathematical representation. Research in 2D intensity image face recognition falls into two categories:

- feature based methods
- holistic (global) methods

Overview of the prominent human face recognition techniques for frontal faces is given below:

2.2 Feature Based Methods

Feature-based approaches reduce the facial image into a vector of geometric feature by identifying, extracting and measuring the unique facial features. Geometric relationships are computed among those points which are further used to match faces using standard statistical pattern recognition techniques.

1. Advantages

Since extraction of the geometric features is performed before matching the test image to training ones, therefore, they are robust to position variations in the input test or training image [5]. They can be made invariant, in principle, to lighting, size or orientation [6]. Also such schemes perform high speed matching and show a compact representation of the face images [7].

2. Disadvantages

The major difficulty and disadvantage of these techniques is making arbitrary decisions about which features are important [8]. Further processing cannot compensate for the intrinsic deficiency if the feature set lacks discrimination ability [6].

3. Major Techniques

The first semi-automated system was developed in 1960s for face recognition. To automate the process Lesk, Harmon and Goldstein used 21 specific subjective markers e.g. lip thickness, hair color, etc in the 1970s.

The method proposed by Wiskott et al known as elastic bunch graph matching is another prominent approach. Recent variations of this method replace the Gabor features by Histograms of Oriented Gradients (HOGs) and a graph matching strategy.

2.3 Holistic (Global) Methods

Holistic approaches perform operations on whole face image (global representation) instead of local representation.

1 Advantages

Unlike feature based approaches, they don't eradicate any details but instead utilize the whole face, therefore, providing with more accurate recognition results.

2 Disadvantages

Such techniques mostly require the face to be either surrounded by a simple background or segmented. Also they are sensitive to variations in scale, position, etc. Furthermore, for accurate recognition the input images need to be well-illuminated and roughly frontal. Reason is algorithms depend on quasi-linear or linear analysis and performance is badly affected under background clutter, 3D orientation changes and non-linear illumination variation.

3 *Major Techniques*

Notable holistic feature spaces include Laplacianfaces [11], their variants [12, 13, 14, 15], Fisher's Linear Discriminant (FLD, LDA or Fisherfaces) [10], Principal Component Analysis (PCA or Eigenfaces) [9], Nearest Subspace (NS) [17] and Nearest Neighbor (NN) [16].

2.3.1 Principle Component Analysis (PCA)

FR problem was solved using a standard linear algebra technique namely PCA in 1988 by Sirovich and Kirby. It was a milestone in this field as it required at most one hundred values to precisely code an aligned and normalized image.

1 *Line Fitting and PCA*

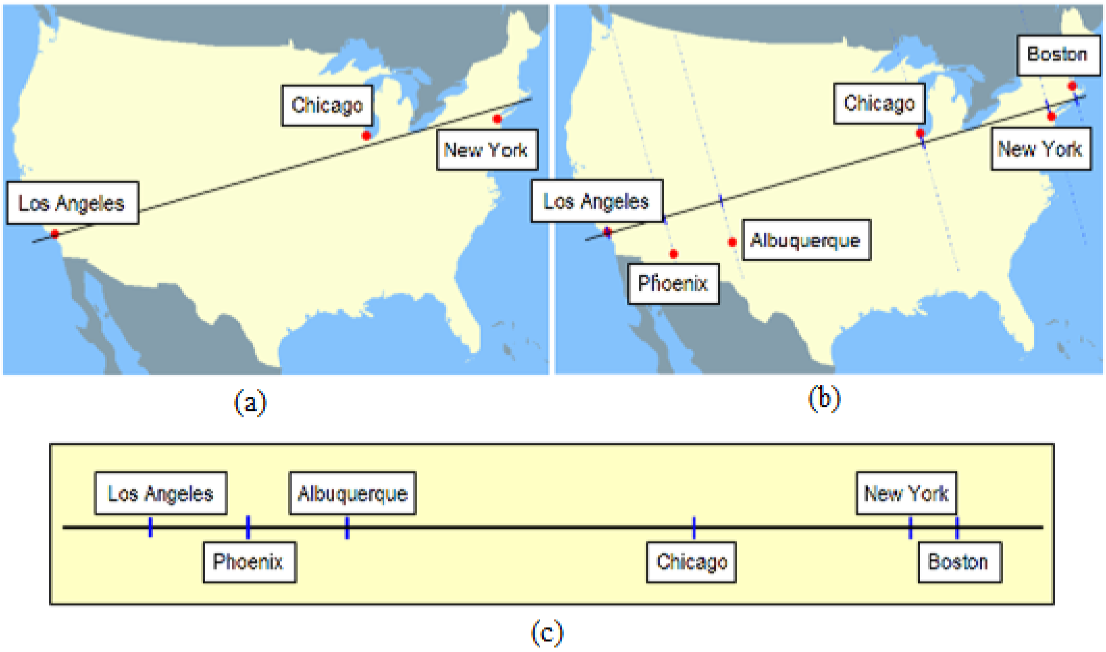

Figure 4: (a) fitting a line to three points. (b) projection of points from the 2D map onto the 1D line (c) distances between points in 1D subspace

Consider a special case called a "least squares line fit" to understand what PCA does [4]. A line has only one dimension, so to reduce 2D to 1D, replace the 2D locations of points with that on a single line. Fig. 4 (a) summarizes the process of how to fit a line on a 2D map with three points. Sum of square of the distance measure from each point location (on 2D) to the line gives the error in the fitting the line. The line which gives the smallest error would be the best-fit line.

2 *Defining a Subspace*

The line in Fig. 4 (a) is a 1D object and located inside 2D space. Therefore, it has a slope/orientation. The slope gives the direction of the line i.e. it computes the in which direction the three points or cities having the maximum spread. The equation of line is given as:

$$y = mx \tag{2}$$

where slope of the line: $m = \frac{dy}{dx}$

The line is, inshort, a *subspace* of the 2D space and description above highlights the direction of line which can separate the points to maximum extent.

3 *Projecting Data Onto a Subspace*

For projecting a point onto a subspace (e.g. convert 2D points to 1D in this case), the location of the subspace closest to its original location in the higher dimensional space is assigned to it.

For example, Fig. 4 shows location of the three cities indicated by blue tic marks that defined 1D subspace (line).

4 *The PCA Subspace*

The first principal component of a given dataset is thus, the direction of maximum separation (e.g. the line itself). The second principal component is the direction of next largest separation i.e. one perpendicular to first principal component. At the most two principal components are present in a 2D dataset.

Many principal components are usually present in an image dataset, as images have large dimension.

Each face image with dimension 40x40 represents just one data point (in a 1,600 dimensional space), in eigenface. Therefore, total number of principal components will always be equal to the number of face images minus one.

5 *Eigenface*

One of the most extensively studied approach of FR is Eigenface.

Turk and Pentland, in 1991, discovered a real-time and reliable automated FR technique. According to this technique, the faces in images can be detected using the residual error.

More formally, considering a set of *N* sample images $\{x_1, x_2, \ldots, x_N\}$ taking values in an *n*-dimensional image space, and assume that each image belongs to one of *c* classes $\{X_1, X_2, \ldots, X_C\}$. Assume a linear transformation mapping the original *n*-dimensional image space into an *m*-dimensional feature space, where $m < n$. The new feature vectors $y_k \in \mathbb{R}^m$ are defined by the following linear transformation:

$$y_k = A^T x_k \qquad k = 1,2, \ldots, N \tag{3}$$

where $A \in \mathbb{R}^{n \times m}$ is a matrix with orthonormal columns.

If the total scatter matrix S_T is defined as

$$S_T = \sum_{k=1}^{N} (x_k - \mu)(x_k - \mu)^T \tag{4}$$

where *N* is the number of sample images, and $\mu \in \mathbb{R}^N$ is the mean image of all samples, then after applying the linear transformation W^T, the scatter of the transformed feature vectors $\{y_1, y_2, \ldots, y_N\}$ is $W^T S_T W$. In PCA, the projection W_{opt} is chosen to maximize the determinant of the total scatter matrix of the projected samples, i.e.,

$$W_{opt} = \arg\ max\, |W^T S_T W| \tag{5}$$

$$= [w_1\ w_2 \ldots w_m]$$

where $\{w_i | i = 1, 2, \ldots, m\}$ is the set of *n*-dimensional eigenvectors of S_T corresponding to the *m* largest eigenvalues. These eigenvectors are called "Eigenpictures" or "Eigenfaces" as they have the same dimension as the original images. If a nearest neighbor classifier is used to preform classification in the reduced feature space and *m* is selected as *N (*the number of images in the training set), then the Eigenface method is equivalent to the correlation.

2.3.2 Linear Discriminant Analysis (LDA)

Fisher Discriminant Criterion (Fisher 1936) finds a small number of features that can differentiate between subject faces but recognizes faces of the same subject.

Linear Discriminant Analysis (FLD or LDA) [10] tries to make the scatter more suitable for classification by reshaping, therefore, it's a *class specific method*. This method selects *W* in [1] in such a way that the ratio of the between-class scatter and the within-class scatter is maximized.

Let the between-class scatter matrix be defined as

$$S_B = \sum_{i=1}^{c} N_i (\mu_i - \mu)(\mu_i - \mu)^T \tag{6}$$

and the within-class scatter matrix be defined as

$$S_W = \sum_{i=1}^{c} \sum_{x_k \in X_i} (x_k - \mu_i)(x_k - \mu_i)^T \tag{7}$$

where μ_i is the mean image of class X_i , and N_i is the number of samples in class X_i. If S_W is nonsingular, the optimal projection W_{opt} is chosen as the matrix with orthonormal columns which maximizes the ratio of the determinant of the between-class scatter matrix of the projected samples to the determinant of the within-class scatter matrix of the projected samples, i.e.,

$$W_{opt} = \arg\max_{\mathrm{W}} \frac{|W^T S_B W|}{|W^T S_W W|} \tag{8}$$

$$= \quad [w_1 \; w_2 \ldots \; w_m]$$

where $\{w_i | i = 1, 2, \ldots, m\}$ is the set of generalized eigenvectors of S_B and S_W corresponding to the *m* largest generalized eigenvalues $\{\lambda_i | i = 1, 2, \ldots, m\}$, i.e.,

$$S_B w_i = \lambda_i \, S_W w_i, \qquad i = 1, 2, \ldots, m \tag{9}$$

Note that there are at most *c* - 1 nonzero generalized eigenvalues, and so an upper bound on *m* is *C* - 1, where *C* is the number of classes.

2.3.3 Nearest Neighbor (NN)

The Nearest Neighbor is a type of compressive classifiers and was used for classification under Random Projection (RP) dimensionality reduction in [1]. Its classification criteria depends on distance measure between the test sample and each training sample. Two popular distance measures are given below:

Euclidean distance $\left(\left\|v_{i,test} - v_{i,j}\right\|_2, i = 1 \ldots C \; and \; j = 1, \ldots, n_i\right)$

Cosine distance $\left(\langle v_{i,test}, v_{i,j}\rangle, i = 1 \dots C \ and \ j = 1, \dots, n_i\right)$

where $v_{i,test}$ is the test sample belonging to the ith class and $v_{i,j}'s$ are the training samples of the ith class.

2.3.4 Nearest Subspace (NS)

The Nearest Subspace [18] assumes samples from each class lie on a hyper-plane specific to that class. According to this assumption, the training samples of a particular class span a subspace. Thus the problem of classification is to find the correct hyperplane for the test sample. According to this assumption, any new test sample belonging to that class can thus be represented as a linear combination of the test samples, i.e.

$$v_{k,test} = \sum_{i=1}^{n_k} \alpha_{k,i} \cdot v_{k,i} + \varepsilon_k \tag{10}$$

where $v_{k,test}$ is the test sample (i.e. the vector of features) assumed to belong to the kth class, $v_{k,i}$ is the ith training sample of the kth class, and ε_k is the approximation error for the kth class. Owing to the error term in equation (10), the relation holds for all the classes *k=1…C.* In such a situation, it is reasonable to assume that for the correct class the test sample has the minimum error ε_k.

To find the class that has the minimum error in equation (10), the coefficients $\alpha_{k,i}$, *k=1…C* must be estimated first. This can be performed by rewriting (10) in matrix-vector notation

$$v_{k,test} = V_k \alpha_k + \varepsilon_k \tag{11}$$

Where $V_k = [v_{k,1}|v_{k,2}| \dots |v_{k,n_k}]$ and $\alpha_k = [\alpha_{k,1}, \alpha_{k,2} \dots \alpha_{k,n_k}]^T$

The solution to (11) can be obtained by minimizing

$$\hat{\alpha}_k = \arg\min_{\mathrm{a}} \left\| v_{k,test} - V_k \alpha \right\|_2^2 \tag{12}$$

The matrix V_k may be underdetermined, i.e. the number the number of samples may be greater than the dimensionality of the inputs. In such a case, instead of solving (12), Tikhonov regularization is employed so that the following is minimized

$$\hat{\alpha}_k = \arg\min_{\mathrm{a}} \left\| v_{k,test} - V_k \alpha \right\|_2^2 + \lambda \|\alpha\|_2^2 \tag{13}$$

The analytical solution of (13) is

$$\hat{\alpha}_k = \left({V_k}^T V_k + \lambda I\right)^{-1} {V_k}^T v_{k,test} \tag{14}$$

Plugging this expression in (11), and solving the error term, we get

$$\varepsilon_k = \left(\left({V_k}^T V_k + \lambda I\right)^{-1} {V_k}^T - I\right) v_{k,test} \tag{15}$$

Based on equations (11-15), the NSC algorithm has the following steps:

NS Algorithm

- Training
  For each class ‘k’, by computing the orthoprojector (the term in brackets in equation (15)).

- Testing
  Calculate the error for each class ‘k’ by computing the matrix vector product between the orthoprojector and $v_{k,test}$.
  Classify the test sample as the class having the minimum error ($||\varepsilon_k||$).

Chapter 3

Face Recognition via Sparse Representation

The technique, called *Sparse Representation based Classification (SRC)*, gives a new solution for classifying subjects using frontal faces with varying illumination, expression, disguise or occlusion.

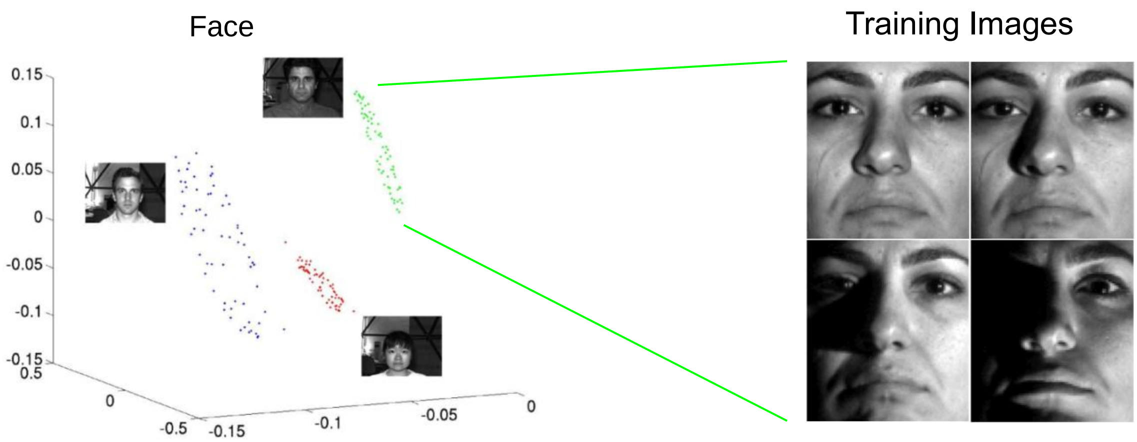

Figure 5: Under varying illumination, images of the same face lie approximately on a lower (nine)-dimensional subspace [19].

The blessing of dimensionality is real data highly concentrate on low-dimensional, sparse, or degenerate structures in the high-dimensional space as shown in Fig. 5. But nothing is free as gross errors and irrelevant measurements are now ubiquitous in massive cheap data. The key challenge is to efficiently and reliably recover sparse or degenerate structures from high-dimensional data, despite gross observation errors. Several characteristics of approximation error or occlusion ε are:

- Randomly supported errors (location is unknown and unpredictable)
- Gross errors (arbitrarily large in magnitude)
- Sparse errors? (concentrated on relatively small part(s) of the image)

SRC is based on the assumption that the training samples of a particular class approximately form a linear basis for a new test sample belonging to the same class. If $\boldsymbol{y}$ is the test sample belonging to the ith class then,

$$v_{k,test} = \alpha_{k,1}v_{k,1}+\alpha_{k,2}v_{k,2}+\cdots+\alpha_{k,n_k}v_{k,n_k}+\ \varepsilon_k \tag{16}$$

$$= \sum_{i=1}^{n_k} \alpha_{k,i} \cdot v_{k,i} + \varepsilon_k$$

where $v_{k,i}$'s are the training samples of the kth class and ε_k is the approximation error (assumed to be Normally distributed).

Equation (16) expresses the assumption in terms of the training samples of a single class. Alternatively, it can be expressed in terms of all the training samples such that

$$y = \alpha_{1,1} + \cdots \alpha_{k,1} v_{k,1} + \cdots + \alpha_{k,n_k} v_{k,n_k} + \alpha_{C,n_C} v_{C,n_C} + \varepsilon$$
$$= \sum_{i=1}^{n_1} \alpha_{1,i} \cdot v_{1,i} + \cdots \sum_{i=1}^{n_k} \alpha_{k,i} \cdot v_{k,i} + \cdots \sum_{i=1}^{n_C} \alpha_{C,i} \cdot v_{C,i} + \varepsilon \tag{17}$$

where C is the total number of classes.

In matrix vector notation, equation (17) can be expressed as

$$y = V\alpha + \varepsilon \tag{18}$$

Where $V = [v_{1,1}| \ldots v_{k,1}| \ldots |v_{k,n_k} \ldots |v_{C,n_C}]$ and $\alpha_k = [\alpha_{1,1} \ldots \alpha_{k,1} \ldots \alpha_{k,n_k} \ldots \alpha_{C,n_C}]^T$

The linearity assumption in [20] coupled with the formulation (18) implies that the coefficients vector α should be non-zero only when they correspond to the correct class of the test sample.

Based on this assumption the following sparse optimization problem was proposed in [20]

$$\arg\min_\alpha \|\alpha\|_1 \text{ subject to} \|y - V\alpha\|_2 \leq \varepsilon \tag{19}$$

The sparse classification (SC) algorithm proposed in [20] is the following:

3.1 Sparse Representation Classifier Algorithm

Table 1: The SRC Algorithm

1. Normalize the columns of $\boldsymbol{V}$ to have unit l_2-norm.
2. Code $\boldsymbol{y}$ over $\boldsymbol{V}$ via l_1-minimization
$$\hat{\alpha}_1 = \arg\min_\alpha \|\alpha\|_1 \text{ subject to} \|y - V\alpha\|_2 \leq \varepsilon \tag{19}$$
where constant ε is to account for the dense small noise in $\boldsymbol{y}$, or to balance the coding error of $\boldsymbol{y}$ and the sparsity of α.
3. Compute the residuals
$$r_i(y) = \|y - V_i \hat{\alpha}_i\|_2 \tag{20}$$
where $\hat{\alpha}_i$ is the coding coefficient vector associated with class i.
4. Output the identity of $\boldsymbol{y}$ as
$$identity(y) = argmin_i\{r_i\}$$ (2

The main step in SRC algorithm is the optimization problem (18) while the rest are straightforward.

Although the pseudoinverse of A is another solution to this optimization problem but $\hat{\alpha}_2$in this case doesn't provide sufficient information for recognizing the test sample y. Fig. 6 shows $\hat{\alpha}_1$ is generally dense, with large nonzero entries corresponding to training samples from many different classes.

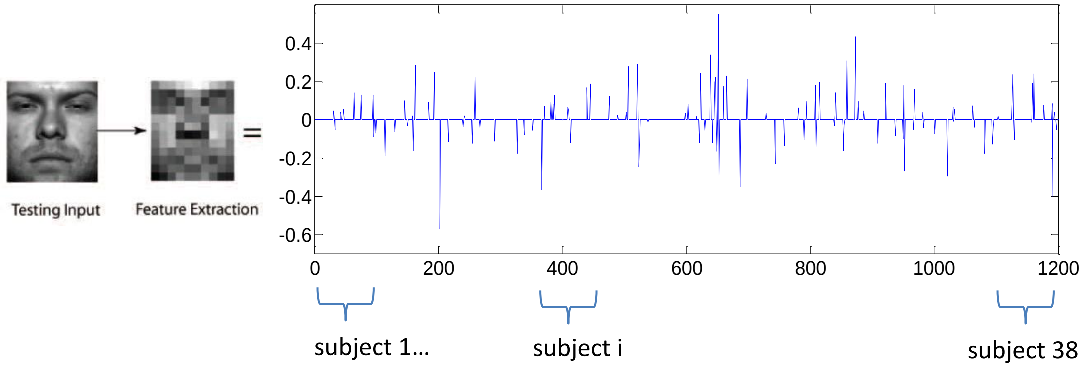

Figure 6: Coefficients of test image from Extended Yale B using l^2-minimization. Large coefficients do not correspond to training images of test subject 1.

3.2 Strengths of SRC

It's one of the most thoroughly studied approach of recent times. The key advantages of this technique, according to the authors are its accuracy and robustness. It targets the two fundamental issues:

- Choice/selection of features
- Robustness to Occlusion and Disguise

The authors claim that unlike other approaches, its performance remains stable, independent of the selected feature set as can be seen in the graph in Fig. 7.

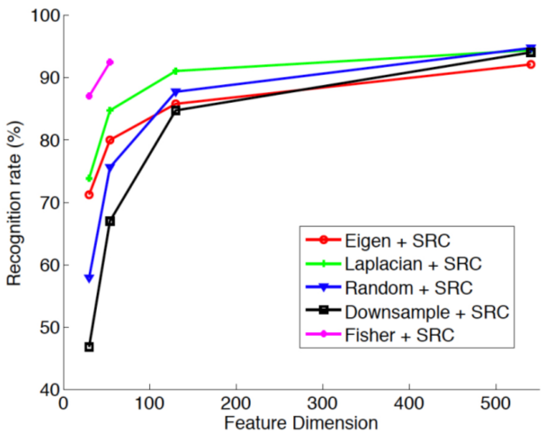

Figure 7: Recognition rates of SRC for various feature transformations and dimensions [20]. The training and query images are selected from the public AR database.

Representation of the input image is computed w.r.t. all available training images as a whole. An extra constraint imposed by the method is smallest number of training images should be used for optimal representation. Fig. 8 shows a sparse representation where majority of the coefficients are zero.

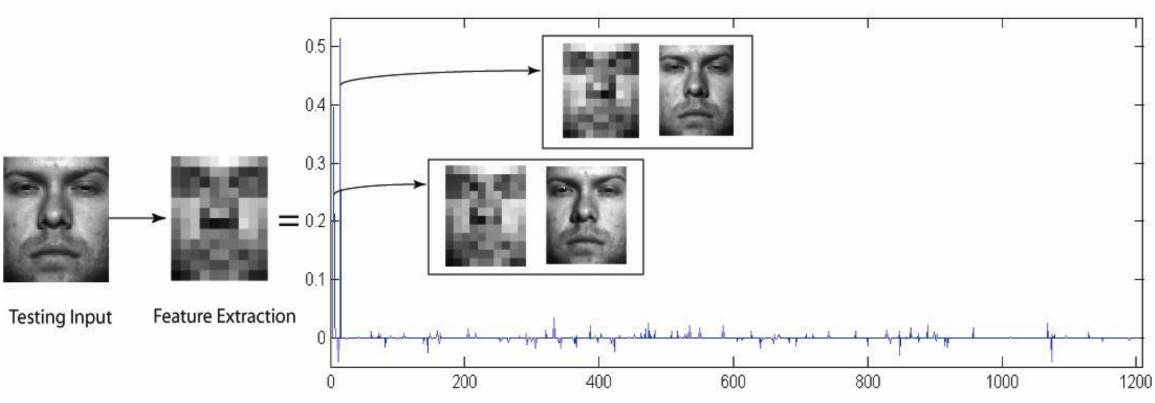

Figure 8: Sparse representation (with downsampled images 12x10) of a test image w.r.t. 1200 training images. Large coefficients correspond to the training images with the same identity as the input image.

For face recognition, facial disguise and image distortion are a great challenge. A major advantage of this technique is that the problem of image occlusion/distortion along with face recognition can be handled under the same framework. All the image distortions shown in Fig. 9 can be dealt while maintaining high accuracy.

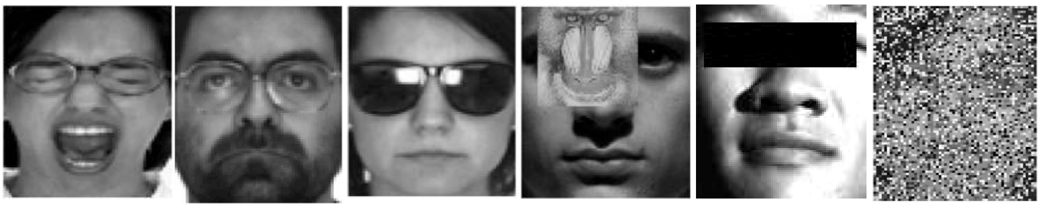

Figure 9: Examples of image distortion on face images.

3.3 Weakness of SRC

The major weak point of this technique is the complexity level and time consumption because of the expensive l1-minimization technique. We overcome this difficulty by using Collaborative Representation [21] instead of sparse representation.

Chapter 4

Face Recognition via Collaborative Representation

There are five representative fast l1-minimization approaches: Shrinkage-Thresholding, Augmented Lagrange Multiplier (ALM), Gradient Projection, Homotopy and Iterative Proximal Gradient [40]. For face recognition, Homotopy [44], ALM and l1_ls [45] are better for their fast speed and accuracy while first order l1-minimization techniques (e.g., ALM [43], FISTA [42] and SpaRSA [41]) are efficient in case of noisy data.

The working mechanism of SRC has not been completely understood, although it's a widely studied technique. For face classification in literature, major emphasis is laid on the role of l1-norm sparsity while the role of collaborative representation (CR), i.e., using the training samples from all classes to represent the query sample y, is much ignored. Sparsity based classification schemes such as SRC become very expensive due to l1-minimization. Recently some researchers have questioned the use of l1-norm sparsity for FR [46], [47].

In order to collaboratively represent the query sample using V with low computational burden, author proposed to use the regularized least square method.

$$\alpha_i = \arg\min_\alpha \{\|y - V_i\alpha\|_2^2 + \lambda\|\alpha\|_1\} \tag{22}$$

where λ is the regularization parameter. Both the representation error $e_i = \|y - V_i\alpha_i\|_2$ and the sparsity term $\|\alpha_i\|_1$ can be used to classify the sample. The l2-norm can also be used to regularize α_i, and the l1-norm and l2-norm lead to almost the same result.

The regularization term gives two advantages. First, it makes the least square solution stable, and second, it introduces a certain amount of "sparsity" to the solution $\hat{\alpha}$, yet this sparsity is much weaker than that by l1-norm.

The solution of CR with regularized least square in Eq. (22) can be easily and analytically derived as

$$\hat{\alpha} = (\mathrm{V}^T\mathrm{V} + \lambda.I)^{-1}\mathrm{V}^T y \tag{23}$$

Let $A = (\mathrm{V}^T\mathrm{V} + \lambda.I)^{-1}\mathrm{V}^T$. Clearly, A is independent of y so that it can be pre-calculated as a projection matrix. Once a query sample $\boldsymbol{y}$ comes, it can simply be projected $\boldsymbol{y}$ onto $\boldsymbol{A}$ via $\boldsymbol{Ay}$. This makes CR very fast.

In addition to the class specific representation residual $\|y - V\alpha\|_2$, where $\hat{\alpha}_i$ is the coefficient vector associated with class i, the l_2-norm "sparsity" $\|\hat{\alpha}_i\|_2$ can also bring some discrimination information for classification. Therefore both of them are used in classification.

4.1 Collaborative Representation Classifier Algorithm

Table 2: The CRC Algorithm
1. Normalize the columns of $\boldsymbol{V}$ to have unit l_2-norm.
2. Code $\boldsymbol{y}$ over $\boldsymbol{V}$ by $$\hat{\alpha} = Ay$$ where $A = (\mathrm{V}^T\mathrm{V} + \lambda.I)^{-1}\mathrm{V}^T$.
3. Compute the regularized residuals $$r_i(y) = \|y - V_i\hat{\alpha}_i\|_2 / \|\hat{\alpha}_i\|_2 \quad (24)$$
4. Output the identity of $\boldsymbol{y}$ as $$identity(y) = argmin_i\{r_i\}$$

4.2 Strengths of CRC

- CRC_RLS perfectly replicates SRC results for image classification while greatly reducing the complexity level.
- It showed why discrimination is improved by sparsity.
- CR plays the major role for classification in SRC and not l1-minimization
- $A = (\mathrm{V}^T\mathrm{V} + \lambda.I)^{-1}\mathrm{V}^T$ is computed offline thus its more time saving and remarkably efficient for online comparison to test subjects

Chapter 5

Face Recognition via Robust NCC

5.1 Traditional NCC

Normalized Cross-Correlation Coefficient (NCC) is a popular measure for template matching if linear changes between the reference and the template images are expected. NCC between two face images V and y is defined as

$$NCC = \frac{\sum (V - \bar{V})(y - \bar{y})}{\sqrt{\sum (V - \bar{V})^2 \sum (y - \bar{y})^2}} \tag{25}$$

Where $\bar{V}$ and $\bar{y}$ are their respective arithmetic means. NCC is computed between the test image and every training image. The class of test image is one whose training image bears maximum NCC.

1 Strong points

- invariant to linear brightness variations
- confined in the range between –1 and 1 (where 1 represents perfect match, 0 represents no match, -1 represents perfect match but complemented)
- similarity doesn't change if an arbitrary (non-negative) number is added or multiplied to every pixel

2 Weak points

NCC doesn’t have the ability to deal with bad pixels. It requires that all the corresponding pixels of the test and training image should follow the same linear transformation. The partial variations in the test and/or training face of the same class destroy this expectation due to expression, occlusion and other variations. It should be noted that randomly located nonconforming pixels are equally bad for NCC based matching as are the coherent pixels. This leads to lowering the peak value with the actual class images and the test image is matched with the wrong class. We shall show in simulations section how the accuracy of NCC is low due to variations in the compared images.

Also, speed requirements are not met for time-critical applications by the traditional normalized cross-correlation due to series of multiplication operations. Thus, this process is time consuming.

5.2 Robust NCC (RNCC)

Our proposed method is called Robust NCC because it overcomes the shortcomings of traditional NCC.

There are thirteen renowned ways to implement the Correlation coefficient [22]. Each one of these has a unique objective. We selected “Correlation as a Rescaled Variance of the Difference between Standardized Scores” and extended it for removing bad pixels from the desired image as NCC in its standard form of Equation (25) doesn’t have this ability.

One variant of NCC is defined as [22]

$$NCC = 1 - 0.5 \sum (\tilde{V} - \tilde{y})^2 \tag{26}$$

Here $\tilde{V}$ and $\tilde{y}$ are zero-mean unit norm vectors representing training and test images, respectively. The term on the right hand side represents sum of squared differences (dissimilarity) between corresponding pixels of training and test images. We notice that the pixels that decrease NCC to a larger extent are represented by the large squared difference values.

As in case of PCA, the variation due to lighting is greatly reduced by discarding the three most significant principal components. We also suggest a similar approach to sort in descending order the squared differences and take a trimmed summation whereby excluding differences corresponding to a certain percentage of pixels from both extremes. Trimming on the higher side is justified in Fig. 10. Small pixels values and a darker image like Fig. 10(c), is obtained after taking the squared difference when the test image matches the correct class. In case of incorrect class, large pixel values and a brighter image is obtained.

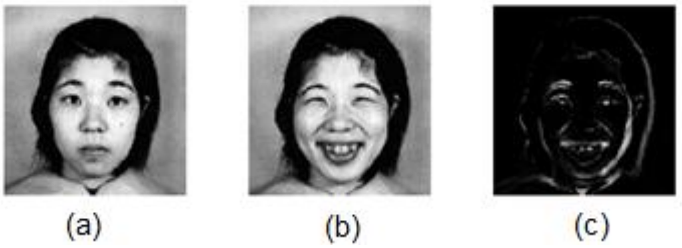

Figure 10: (a) Normal Face (b) Face with expression variation (c) The third image shows squared differences at the corresponding pixels. We see that large values are mostly at the expression pixels i.e., eyes brows, eyes, lips and teeth only.

Trimming on lower side helps to keep summation high for non-matching faces. Our scheme is based on estimating robust correlation via the rejection of outliers on both sides.

$$RNCC = 1 - 0.5\sum_{i=L}^{H}[(\tilde{V} - \tilde{y})^2]_{sort} \tag{27}$$

Fig. 11 shows how the variation in trim affects the Recognition rate in case of AR database. We have used a trimming of 10% values from the higher and 35% values from lower side, giving rise to H and L being equal to 0.1d and 0.65d, respectively. Here d is the dimension of the faces being matched. Small improvement in accuracy can be achieved by varying this clip ratio but we have kept this clip ratio same for all databases.

Trimming	1	5	10	15	20	25	30	35	40	45	50
0	0.65	0.78	0.80	0.80	0.79	0.79	0.78	0.78	0.77	0.76	0.75
5	0.65	0.78	0.80	0.80	0.79	0.79	0.78	0.78	0.77	0.76	0.75
10	0.65	0.78	0.80	0.80	0.79	0.79	0.78	0.78	0.77	0.76	0.75
15	0.65	0.78	0.80	0.80	0.79	0.79	0.78	0.78	0.77	0.76	0.76
20	0.65	0.78	0.80	0.80	0.79	0.79	0.78	0.78	0.77	0.76	0.75
25	0.65	0.77	0.80	0.80	0.79	0.79	0.78	0.78	0.77	0.76	0.76
30	0.65	0.77	0.80	0.80	0.79	0.79	0.79	0.78	0.77	0.77	0.76
35	0.65	0.77	**0.80**	0.80	0.80	0.79	0.79	0.78	0.77	0.76	0.75
40	0.65	0.77	0.80	0.80	0.80	0.79	0.79	0.78	0.78	0.76	0.76
45	0.65	0.77	0.80	0.80	0.79	0.79	0.79	0.78	0.78	0.76	0.76
50	0.64	0.77	0.80	0.80	0.79	0.79	0.80	0.78	0.78	0.77	0.00

Figure 11: Trimming variation on AR database

Selected Trim = (10 : 65)%

The ultimate goal of this trimming is to remove the outliers claiming match in case of few very large squared differences and also claiming mismatch in case of few very small squared differences. Rather we choose the mediocre ones for a more refined match.

Note that the term $\tilde{V}$ can be computed offline for all training images. We down sample training and test images to $m \times n$ size. The complexity is dominated by mn squares and their sorting. Fig. 14 shows a successful example from the AR database. Notice that trimming compensates for small misalignment of the image edges, as well as occlusion due to sunglasses.

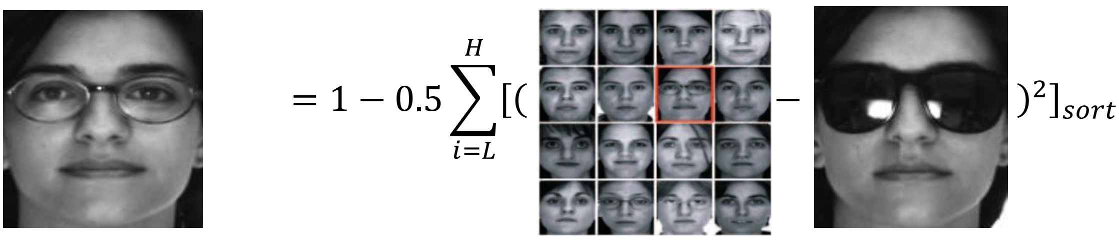

Figure 12: Overview of our approach. Our algorithm determines the true identity (indicated with a red box at second row and third column) of an occluded test image from 700 training images of 100 individuals (7 each) in the standard AR face database**.**

1 *Strong points*

- Outperform specifically in case of expression [24] and occlusion
- Easily deals with bad pixels/outliers
- Reduced Complexity than traditional NCC
- As fast as CRC_RLS
- Histogram equalization is a better option instead of mean subtraction for illumination variation as will be proved by experimental verification

2 *Weak points*

- Not so effective for handling extreme cases of illumination variation even after mean subtraction

Chapter 6

Face Recognition despite Occlusion and Disguise

In order to test RNCC's ability to handle real malicious occlusions, we choose a subset of the AR Face database containing 1,399 images (14 each, except for a corrupted image w-027-14.bmp) of 100 subjects, 50 male and 50 female. For training, we use 799 images (about 8 per subject) of unoccluded frontal views with varying facial expression. The images are resized to 83×60, so in this case, $V \in \mathbb{R}^{4980\times799}$ matrix.

We consider two separate test sets of 200 images. The first test set contains images of the subjects wearing sunglasses, which occlude roughly 20 percent of the image. The second test set considered contains images of the subjects wearing a scarf, which occludes roughly 40 percent of the image.

Figure 13: RNCC failed with the whole image (holistic) for occlusion with scarf and glasses.

Fig. 13 shows one such failure where the test image(right) is wearing scarf and glasses, which occludes roughly 70 percent of the image. Notice that the largest coefficient correspond to an image of bearded man (top left) whose mouth region resembles the scarf and eyes region is occluded by glasses. The second and third coefficients (bottom left) also correspond to bearded men.

Table 3: Table of recognition rates for occlusion on the AR database

Algorithms	Rec. rate sunglasses	Rec. rate scarves
SRC	87.0%	59.5%

(partitioned)	**(97.5%)**	**(93.5%)**
RNCC (partitioned)	78.5% **(98%)**	8.0% **(93.5%)**

Table 3 compares RNCC to SRC [20]. Without partitioning SRC achieves a recognition rate of 87 percent (top left), more than 8.5 percent better than RNCC. For occlusion by scarves, its recognition rate is 59.5 percent (top right), almost six times better than RNCC but still quite poor. To overcome occlusion, authors of SRC suggest partitioning each image into multiple smaller blocks, applying SRC algorithm to each of the blocks and then find the correct match by aggregating the results through voting.

Partitioning increases the recognition rate on scarves from 10.5 percent to 95 percent and also improves the recognition rate on sunglasses from 78.5 percent to 99 percent. The recognition rates are 1.5% better than SRC in case of scarves and sunglasses. This performance exceeds the best known results to date on the AR data set [25], [20].

A major advantage of this technique is that the problem of image occlusion/distortion along with face recognition can be handled under the same framework just like SRC but also with much less complexity.

The major differences in SRC and RNCC when used for occlusion are summarized in table below for 1 test image:

Table 4: Comparison for occlusion using 1 test and 1 training image

SRC	**RNCC**
runs for 4×2 times and then performs voting.	only once
Increases training matrix size from A$\in \mathbb{R}^{D\times n}$ to $B \in \mathbb{R}^{D\times(D+n)}$ e.g. $B = [A, I] \in \mathbb{R}^{4980\times(4980+799)}$	Doesn't increase training matrix size
2.27GHz machine with 2GB RAM is out of memory Claim: takes 75s on PowerMac G5	3 seconds

Contrary to [20], we suggest to keep the training matrix size to its original dimensions. Without adding any further complexity to it (i.e. increasing the training matrix size), we aggregate all the blocks of all the images and apply RNCC just once for one test image comparison.

6.1 Improving Recognition by Block Partitioning and Weighted Patch Matching

We partition each of the training images into L blocks of size $a \times b$, producing a set of matrices $V^{(1)}, \dots, V^{(L)} \in \mathbb{R}^{p\times m}$, where $p \doteq ab$. We similarly partition the test image y into $y^{(1)}, \dots, y^{(L)} \in \mathbb{R}^{p}$. Fig. 14 illustrates this scheme.

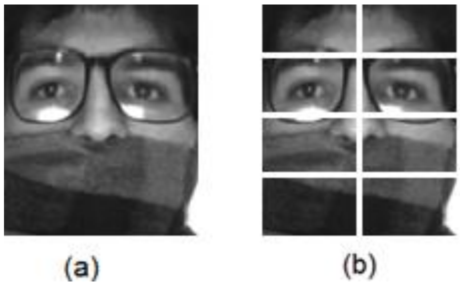

Figure 14: (a) Holistic (b) Partitioned into (4×2) blocks

Contrary to [20] which applies SRC classifier algorithm within each block and is extremely time expensive, we concatenate all the partitioned images row wise forming $B \in \mathbb{R}^{m \times n}$ which has $(a \times b) \times n\ blocks$ for n training images.

1 Steps for Weighted Patch Matching Using RNCC

All the steps after that are same as given in section 5.2, the only difference is step 4 mentioned below. After summing the squared differences we multiply the results of left half face i.e. 4 left most blocks. Reason being if any of these blocks has good correlation with test image, its result will dominate others. Repeat the same process with right half face of each training image.

- Compute $(B-y)^2$ and resize all the $a \times b$ pixel blocks to vectors of size $p \times 1$ pixels where $p \doteq ab$.
- Perform sorting and trimming to remove the outliers as done previously in RNCC classifier algorithm.
- Compute $\sum_{i=L}^{H}[(B-y)^2]_{sort}$ thus summing all the vector blocks resulting into singular values of 1×1 pixel blocks.
- Multiply the $(a \times b) \times n$ singular value blocks row wise of each partitioned image (representative of half face) thus reducing the matrix size to $(1 \times b) \times n$ pixel blocks.
- Sum the two columns (representative of left and right half face) containing singular values of each partitioned image to $1 \times n$ blocks.
- Finally compute $RNCC = 1 - 0.5 \sum_{i=L}^{H}[(B-y)^2]_{sort}$

2 Example of Weighted Patch Matching Using RNCC

The efficacy of the proposed scheme is verified using AR database for faces disguised with sunglasses or scarves. We partition the images into eight (4×2) blocks of size 20×30 pixels. Thus $B \in \mathbb{R}^{4800 \times 799}$.

- Compute $(B-y)^2$ and resize all the 20×30 pixel blocks to vectors of size 600×1 pixels.
- Perform sorting and trimming to remove the outliers as done previously in RNCC classifier algorithm.
- Compute $\sum_{i=L}^{H}[(B-y)^2]_{sort}$ thus summing all the vector blocks resulting into singular values of 1×1 pixel blocks.

- Multiply the $(4 \times 2) \times 799$ singular value blocks row wise of each partitioned image thus reducing the matrix size to $(1\times 2) \times 799$ pixel blocks.
- Finally, sum the two columns containing singular values of each partitioned image to 1×799 blocks.
- Finally compute $RNCC = 1 - 0.5 \sum_{i=L}^{H} [(B - y)^2]_{sort}$

Chapter 7

Face Recognition via Cascaded Classifier RNCR

7.1 Introduction

It is common wisdom that process of decision making improves by gathering a variety of views and inputs. A recent trend has been to combine individual classifiers, since the performance of any classifier is more sensitive to some factors and relatively invariant to others. The goal is to create a system that is more robust than any individual classifier to variables that complicate the recognition task. Such systems have been termed as multiple classifier systems (MCSs) [39] and are a very active research area at present.

7.2 Proposed Confidence based Cascaded Classifier

To further enhance the performance of Robust NCC specifically in case of illumination variation, we propose a confidence based cascaded approach based on the concatenation of the two classifiers, namely Robust NCC and Collaborative Representation. Each classifier produces an estimate of how confident it is in its decision. Confidence of first stage is also useful for reducing the search space for the second classifier. We decide to use RNCC as first stage but any one among the two classifiers can be selected as first stage in order to achieve higher efficiency depending upon a given database with only a minor change in accuracy. For example, it would make system more efficient if CRC is cascaded first in case of Extended Yale B database as RNCC would play lesser role for this illumination variant database.

RNCC produces its Confidence measure by the ratio of the second nearest class and the nearest class score.

$$C1 = 100 \times \left(1 - \frac{\max RNCC\ (2nd\ class)}{\max RNCC\ (1st\ class)}\right) \tag{28}$$

The confidence $C1$ of $RNCC$ is compared with a fixed threshold say 40 which is selected by seeing the behavior of the classifier on a given database.

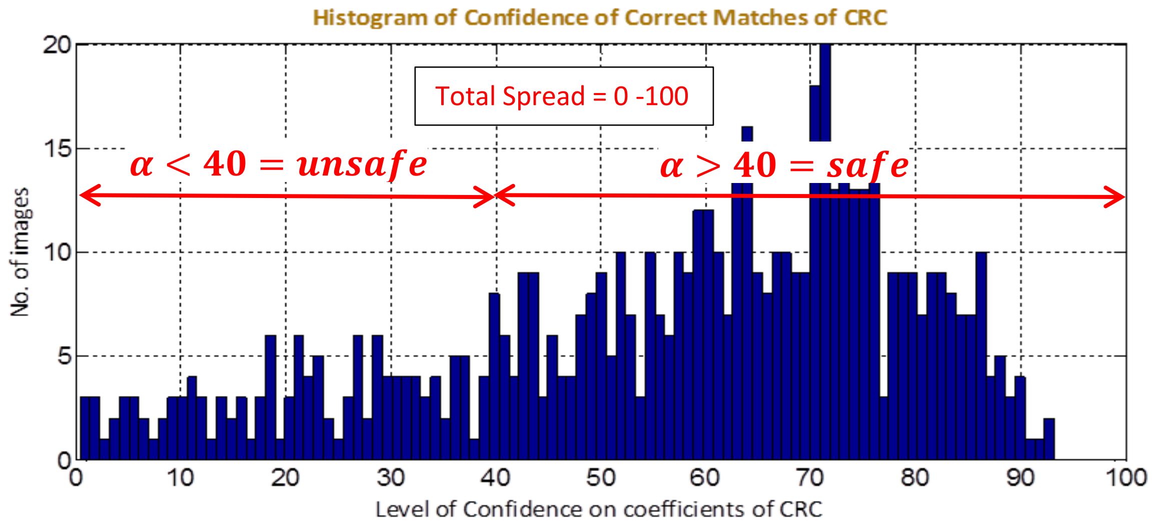

Figure 15: Confidence of CRC for correct matches on AR database

Fig. 15 shows the spread of Level of Confidences of CRC is from 0 to 100 approximately on AR database (dimension =130) with 700 test images (seven images per subject from Session 2 with only expression and illumination variation). The largest peak indicates 20 images having confidence of 72%. Fig. 16 below shows the actual confidence graph of RNCC on AR database.

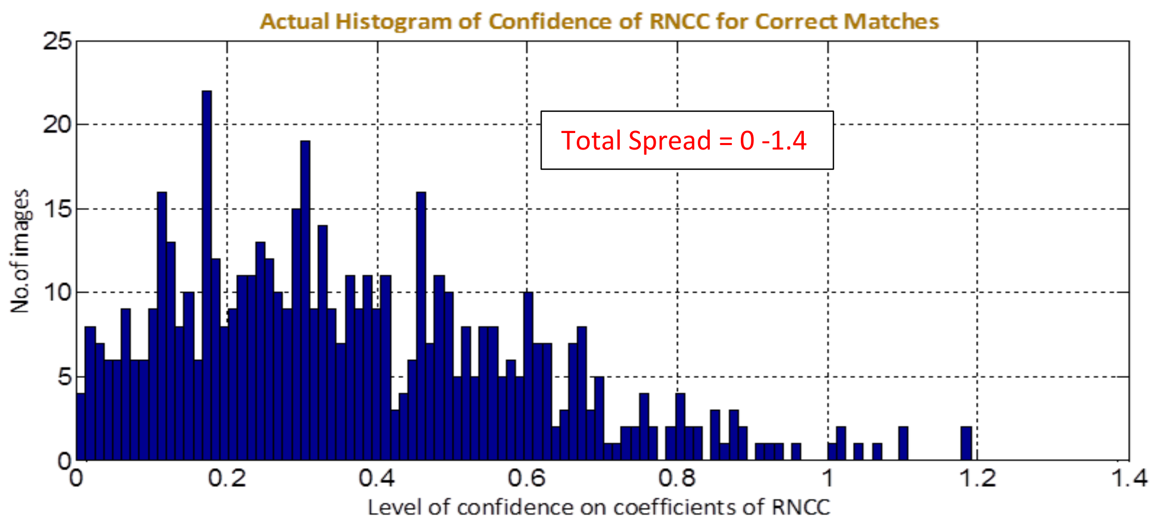

Figure 16: Actual confidence of RNCC for correct matches on AR database

Although the confidence spread is only up to 1.2 but it has the classification power as will be indicated by the results in chapter 8. To compare the confidences of the two algorithms we need to provide a proper shift such that the confidences of RNCC also spread from 0 to 100. We therefore multiply confidences of RNCC with (Total spread/max. confidence level) which in case of AR database is (100/1.3) approximately. A little variation to this selected shift may somewhat improve accuracy but mostly it gives the best results for all the tested databases.

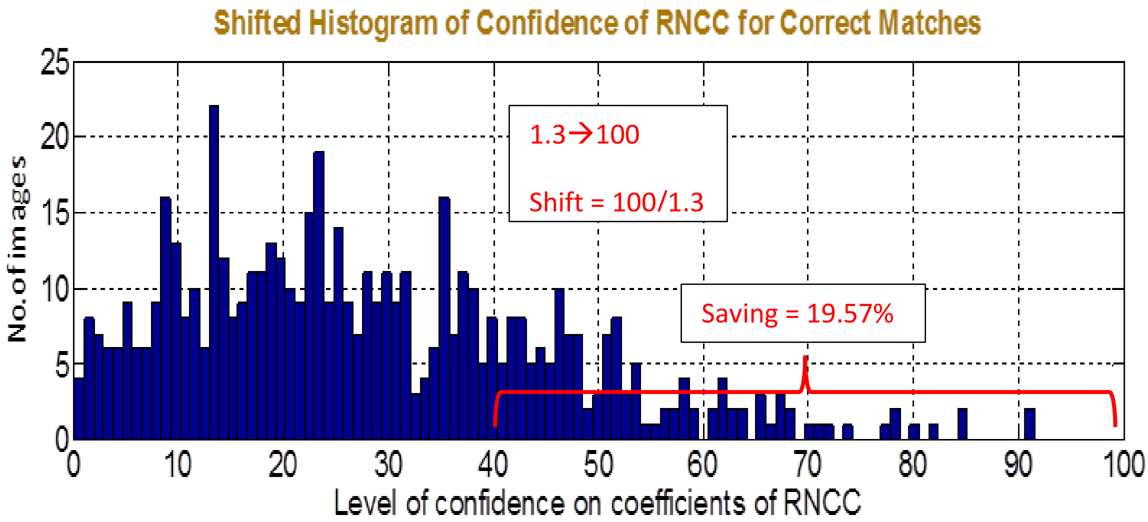

Figure 17: Shifted confidence of RNCC for correct matches on AR database

Fig. 17 shows the shifted confidences of RNCC for correct matches on AR database with saving of 19.57% or 136 images out of 700 which are not sent to second stage.

How much images are sent to second stage varies from database to database. In case of Bosphorus database, saving is up to 96% and only 4% of the images are sent to the second stage due to superior performance of RNCC to expression variation.

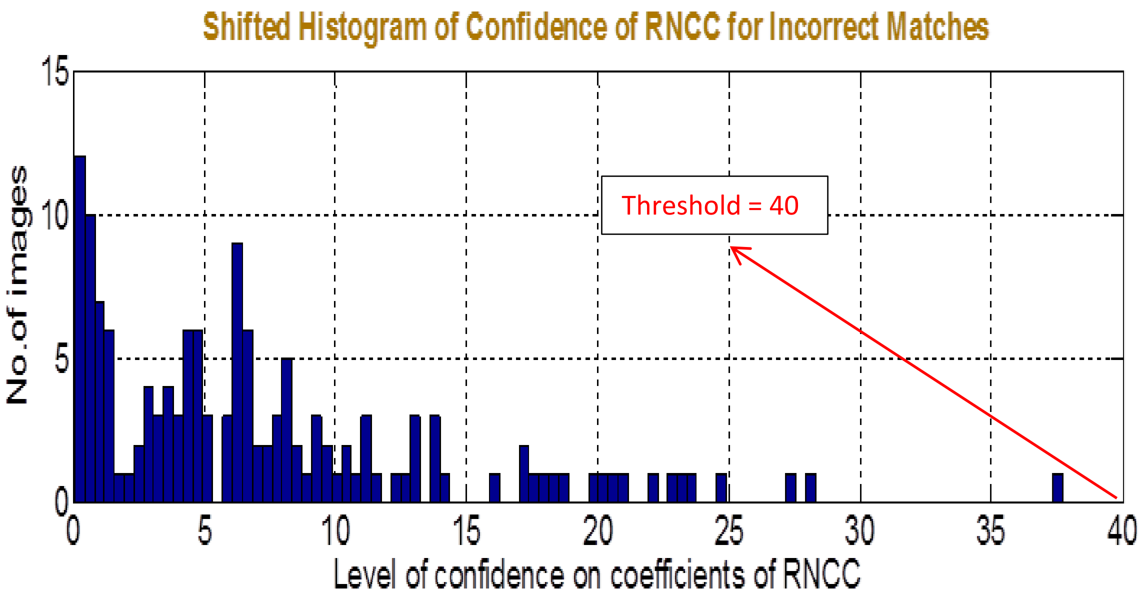

Figure 18: Shifted confidence of RNCC for incorrect matches on AR database

Fig. 18 shows the shifted confidences of RNCC for incorrect matches on AR database which are not more than 40. Therefore a threshold limit of 40 is selected for all databases. For all images with confidence less than threshold limit are sent to 2^{nd} classifer i.e. CRC to classify the test sample and compute the corresponding confidence estimate by the formula below:

$$C2 = 100 \times \left(1 - \frac{\min\ r_i(y)\ (1st\ class)}{\min\ r_i(y)\ (2nd\ class)}\right) \tag{28}$$

7.3 Cascaded Classifier Algorithm

Table 4: The RNCR Algorithm

1. Normalize the columns of $\widetilde{\boldsymbol{V}}$ to have unit l_2-norm where $\widetilde{\boldsymbol{V}}$ is zero-mean unit normalized.
2. Compute

$$RNCC = 1 - 0.5\sum_{i=L}^{H}[(\widetilde{V} - \widetilde{y})^2]_{sort} \quad (27$$

3. Compute Confidence $\boldsymbol{C1}$. If $C1 > threshold$, **Output:**

$$identity(y) = \max \hat{\alpha}(y)$$

4. Else code $\widetilde{\boldsymbol{y}}$ over $\widetilde{\boldsymbol{V}}$ by

$$\hat{\alpha} = A\tilde{y}$$

   where $A = \left(\tilde{V}^T V + \lambda.I\right)^{-1}\tilde{V}^T$.
5. Compute the residuals $r_i(y) = \|y - A_i\hat{x}_i\|_2 / \|\hat{x}_i\|_2$
6. Compute Confidence $\boldsymbol{C2}$. If $(C2 * shift) > C1$, **Output:**

$$identity(y) = \max\ RNCC(y)$$

   Else **Output:**

$$identity(y) = \max \hat{\alpha}(y)$$

1 *Strong points*

- Intelligent cascading due to reduced search space for 2nd stage
- Handles expression, illumination, occlusion better than SRC and CRC_RLS
- More accurate and robust than SRC and CRC_RLS

Chapter 8

Experimental Verification Using Standard Databases

In this section, we present experiments on seven renowned databases for face recognition with illumination, expression, occlusion and disguise to demonstrate the efficacy of the proposed classification algorithm. We tried to keep the training size minimum. Histogram Equalized (h) images instead of zero mean subtraction often helps to improve results specifically in case of illumination variation therefore we have used histogram equalized images for RNCR. Training size is kept minimum.

All the experiments are carried out using MATLAB on a 2.27GHz machine with 2GB RAM. Design parameters in our algorithm are λ = 1e-2, threshold = 40, whereas shift = 100/1.3. Results may be improved by a little variation in the above selected values. (h) and (m) represent histogram equalized images and mean subtracted images respectively as preprocessing steps

Trimming used for all databases is H = 10%, L = 65%. A little variation in the selected trimming ratio often gives better results for certain databases but the stated trimming ratio gives best results for most databases. The images are neither histogram equalized nor mean subtracted in case of SRC and CRC.

For all the test databases, we selected four feature space dimensions: 30, 54, 130, and 540 which correspond to the downsample ratios 1/24, 1/18, 1/12, and 1/6, respectively except Extended Yale B for which feature space dimensions 30, 56, 120, and 504 are specifically selected in order to compare results with SRC [20]. Those numbers correspond to downsampling ratios of 1/32, 1/24, 1/16, and 1/8, respectively.

8.1 Cohn Kanade

The Cohn Kanade database contains 486 short image sequences of 97 different persons performing the six universal facial expressions [30]. Each sequence shows a neutral face at the beginning and then develops into the peak expression.

In [31], image sequences of 62 persons are used which consist of overall 4060 images. 1381 images are used for testing. Maximum face recognition rate achieved was up to 91.17% using binary decision tree (BDT), 98.6% using Bayesian Networks (BN) with 10-fold cross validation.

Using just dimension 30 or (6x5), RNCC achieved 99.61% while for all the higher dimensions its performance was 100%.

8.2 Cohn Kanade Plus

Cohn-Kanade plus Database contains 593 image sequences of 123 subjects. The subjects expressed the 6 basic facial expressions ((i.e., joy, surprise, anger, fear, disgust, and sadness), and up to 23 facial displays.

We use 123 images, 1 neutral image per subject for training and 582 images (peak frames of 6 basic emotions) for testing.100%.

Table 5: Rate of recognition on Cohn Kanade Plus Database

Algorithms	30 (6x5)	54 (9x6)	130 (13x10)	540 (27x20)
CRC	93.15	97.14	99.1	100
RNCC(m)	99.49	100	100	100
RNCC(h)	**100**	100	100	100
RNCR(h)	99.66	**100**	**100**	**100**

As we are the first to use this database, therefore, comparison with other algorithms cannot be performed.

8.3 Bosphorus

This is a very challenging database [29] intended for research on 2D and 3D human face processing tasks. The database consists of 4666 images of 105, 27 of whom were professional actors. The subjects expressed the 6 basic facial expressions and up to 24 AUs. Note that only the 2D frontal face images were used in our experiments with expression variation.

This database is specifically used to test invariance to expression. We use 299 neutral images for training and the rest (Upper face AU (FAU), Lower Face AU (LFAU), Combination AU (CAU), Expression (E)) for testing (2603 images).

Table 6: Rate of recognition on Bosphorus Database

Algorithms	30 (6x5)	54 (9x6)	130 (13x10)	540 (27x20)
CRC	62.35	74.64	88.05	89.7
RNCC(m)	81.71	87.01	90.63	91.89

RNCC(h)	**83.08**	86.75	90.2	90.86
RNCR(h)	73.58	**86.29**	**94.65**	**93.98**

8.4 Jaffe

This Japanese database contains only posed expressions [38]. It contains 213 images of 7 facial expressions (6 basic facial expressions + 1 neutral) posed by 10 Japanese female models. The images are of size 256 × 256.

We use 3 neutral images per subject for training (30 images) and the rest for testing (184 images).

Table 7: Rate of recognition on Jaffe Database

Algorithms	30 (6x5)	54 (9x6)	130 (13x10)	540 (27x20)
CRC	83.7	86.96	90.76	91.85
RNCC(m)	90.76	92.93	95.65	96.2
RNCC(h)	**91.3**	92.39	92.39	95.11
RNCR(h)	86.41	**91.85**	**92.39**	**95.11**

8.5 Yale Face

The Yale Face Database [33] contains 165 images for 15 subjects (11 images each). The images are labeled center-light, w/glasses, happy, left-light, w/no glasses, normal, right-light, sad, sleepy, surprised and wink.

We use 15 images (1 normal per subject) for training and 75 (5 for each subject giving expressions happy, sad, sleep, surprised, wink) for testing. Shift =1 in this case.

Table 8: Rate of recognition on Yale Face Database

Algorithms	30 (6x5)	54 (9x6)	130 (13x10)	540 (27x20)
CRC	89.17	95	99.17	99.17
RNCC(m)	90.67	93.33	96	98.67
RNCC(h)	92	96	98.67	97.33
RNCR(h)	**97.78**	**100**	**100**	**100**

To compare with other algorithms results [34], we use 3 images for training and the rest 8 for testing.

Table 9: RNCR Comparison with Other Algorithms on Yale Face Database	
Algorithms	**150 (15×10)**
RNCR(h)	**100**
CHING-LIANG LU[34]	94.17
Yung-Mao Luo [35]	72.50
2DPCA[36]	61.70
2DFLD[36]	54.20
R. Mutelo et al.[36]	70.80
Wankou Yang[37]	92.50

8.6 Extended Yale B

The Extended Yale B database consists of 2,414 frontal-face images of 38 individuals [26]. For each subject, we randomly select half of the images for training (i.e., about 32 images per subject) and the other half for testing. Randomly choosing the training set ensures that our results and conclusions will not depend on any special choice of the training data.

This database is specifically used to test invariance to illumination. The face recognition rate of RNCR is compared with the renowned technique SRC [20] and RNCC. For illumination variation, as the results clearly show histogram equalized images, RNCC(h) have higher recognition rate as compared to mean subtracted images RNCC(m). For SRC, the recognition rate stated in [20] is used for comparison.

Table 10: Rate of recognition on Extended Yale B Database				
Algorithms	**30 (6x5)**	**56 (8x7)**	**120 (12x10)**	**504 (24x21)**
SRC	74.57	86.16	92.13	97.1
CRC	2.65	89.15	96.02	99.01
RNCC(m)	2.65	54.27	63.96	75.06
RNCC(h)	58.66	72.08	84.18	93.29
RNCR(h)	**79.45**	**94.61**	**98.43**	**99.34**

8.7 AR

The AR database consists of over 4,000 frontal images for 126 individuals (26 images each in two separate sessions) [28]. These images include more facial variations, and facial disguises (scarf and sunglasses). We select a subset of the data set consisting of 50 male subjects and 50 female subjects. For each subject, 14 images with only illumination change and expressions were selected: the seven images from Session 1 for training, and the other seven from Session 2 for testing.

Algorithm RNCR results are compared with SRC [20] and RNCC. The results again indicate a slightly better performance in case of histogram equalized images as database contains images with illumination variation.

Table 11: Rate of recognition on AR Database

Algorithms	30 (6x5)	54 (9x6)	130 (13x10)	540 (27x20)
SRC	46.78	67	84.55	93.85
CRC	57.43	73.43	85.86	92.71
RNCC(m)	61.14	68.71	79.57	84.14
RNCC(h)	63.71	70.43	79.86	86.14
RNCR(h)	**65.86**	**77**	**90**	**95.86**

CONCLUSION

In this paper, we have deduced that when using Normalized Cross Correlation, trimming is critical for the high-performance face recognition under expressions and occlusion. Our simulations indicate that at the most 130 (13 x 10) features are sufficient for good recognition. One can achieve striking recognition performance by intelligently using a simple algorithm such as NCC and cascading.

The corrupted pixels are detected as a byproduct of the NCC calculations keeping the overall computational complexity of the algorithm low. Robustness to occlusion allows the algorithm to tolerate small pose variation or misalignment.

Future work has two directions. The algorithm RNCC is applicable wherever NCC is used under occlusion conditions, e.g., tracking. The second direction is to use rich literature of M and Robust estimators on the squared differences [48].

REFERENCES

[1] G. M. Weiss and J. W. Lockhart, "Identifying User Traits by Mining Smart Phone Accelerometer Data" SensorKDD'11, 2011, San Diego, CA, USA.

[2] New Scientist [Online] [Last Visited: 5 October, 2013]. http://www.newscientist.com/article/mg21528804.200-fbi-launches-1-billion-face-recognition-project.html

[3] K. W. Bowyer, Bowyer, K. Chang, and P. Flynn, "A survey of approaches and challenges in 3D and multi-modal 3D+2D face recognition," in Computer Vision and Image Understanding, 2006, vol. 101, pp. 1–15.

[4] Cognotics – Resources for Cognitive Robotics [Online] [Last Visited: 5 October, 2013]. http://www.cognotics.com/opencv/servo_2007_series/part_4/index.html

[5] T. Jebara, "3D Pose Estimation and Normalization for Face Recognition," Center for Intelligent Machines, McGill University, Undergraduate Thesis May 1996.

[6] I. J. Cox, J. Ghosn, and P. N. Yianilos, "Feature based face recognition using mixture-distance," in Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, 1996, pp.209-216.

[7] R. Brunelli and T. Poggio, "Face Recognition Through Geometrical Features," in Proceedings of the Second European Conference on Computer Vision, Vol.588 Lecture Notes In Computer Science, G. Sandini, Ed. London, UK: Springer-Verlag, 1992, pp.782-800.

[8] R. Cendrillon and B. C. Lowell, "Real-Time Face Recognition using Eigenfaces," in Proceedings of the SPIE International Conference on Visual Communications and Image Processing, Vol.4067, 2000, pp.269-276.

[9] M. Turk and A. Pentland. Eigenfaces for recognition. In Proceedings of the IEEE International Conference on Computer Vision and Pattern Recognition, 1991.

[10]P. Belhumeur, J. Hespanda, and D. Kriegman. Eigenfaces vs. Fisherfaces: recognition using class specific linear projection. IEEE Transactions on Pattern Analysis and Machine Intelligence, 19(7):711–720, 1997.

[11]X. He, S. Yan, Y. Hu, P. Niyogi, and H. Zhang. Face recognition using Laplacianfaces. IEEE Transactions on Pattern Analysis and Machine Intelligence, 27(3):328–340, 2005.

[12]T. Kim and J. Kittler. Locally linear discriminant analysis for multimodally distributed classes for face recognition with a single model image. IEEE Transactions on Pattern Analysis and Machine Intelligence, 27(3):318–327, 2005.

[13]H. Chen, H. Chang, and T. Liu. Local discriminant embedding and its variants. In Proceedings of the IEEE International Conference on Computer Vision and Pattern Recognition, 2005.

[14]S. Yan, D. Xu, B. Zhang, H. Zhang, Q. Yang, and S. Lin. Graph embedding and extension: A general framework for dimensionality reduction. IEEE Transactions on Pattern Analysis and Machine Intelligence, 29:40–51, 2007.

[15]L. Qiao, S. Chen, and X. Tan. Sparsity preserving projections with applications to face recognition. Pattern Recognition, 43(1):331–341, 2010.

[16]R. Duda, P. Hart, and D. Stork, Pattern Classification, second ed. John Wiley & Sons, 2001.

[17] J. Ho, M. Yang, J. Lim, K. Lee, and D. Kriegman, "Clustering Appearances of Objects under Varying Illumination Conditions," Proc. IEEE Int'l Conf. Computer Vision and Pattern Recognition, pp. 11-18, 2003.

[18]A. Majumdar and R. K. Ward, “Nearest Subspace Classifier” submitted to International Conference on Image Processing (ICIP09).
[19]Ronen Basri and David Jacobs, “Lambertian reflectance and linear subspaces,” IEEE Transactions on Pattern Analysis and Machine Intelligence,25(2): 218-233, 2003.
[20]Y. Yang, J. Wright, Y. Ma and S. S. Sastry, “Feature Selection in Face Recognition: A Sparse Representation Perspective”, IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 1 (2), pp. 210-227, 2009.
[21]L. Zhang, M. Yang, and X. Feng, “Sparse representation or collaborative representation: which helps face recognition?” in Proceedings of the IEEE International Conference on Computer Vision (ICCV '11), pp. 471–478, November 2011.
[22]J. L. Rodgers, W. A. Nicewander, “Thirteen Ways to Look at the Correlation Coefficient,” The American Statistician, Vol. 42, No. 1. (1988), pp:59-66.
[23]Black Hat Security Conference [Online] [Last Visited: 5 October, 2013]. http://www.eweek.com/c/a/Security/Facial-Recognition-Tools-Facebook-Photos-Threaten-End-to-Anonymity-Study-565285/
[24]A. Zafar, R. Nawaz, “Face Recognition with expression variation via Robust NCC,” submitted to IEEE International Conference on Emerging Technologies (ICET13).
[25]F. Sanja, D. Skocaj, and A. Leonardis, “Combining Reconstructive and Discriminative Subspace Methods for Robust Classification and Regression by Subsampling,” IEEE Trans. Pattern Analysis and Machine Intelligence, vol. 28, no. 3, Mar. 2006.
[26]A. Georghiades, P. Belhumeur, and D. Kriegman, “From Few to Many: Illumination Cone Models for Face Recognition under Variable Lighting and Pose,” IEEE Trans. Pattern Analysis and Machine Intelligence, vol. 23, no. 6, pp. 643-660, June 2001.
[27]K. Lee, J. Ho, and D. Kriegman, “Acquiring Linear Subspaces for Face Recognition under Variable Lighting,” IEEE Trans. Pattern Analysis and Machine Intelligence, vol. 27, no. 5, pp. 684-698, 2005.
[28]A. Martinez and R. Benavente, “The AR Face Database,” CVC Technical Report 24, 1998.
[29]A Sarvan et al, “Facial Action Unit Detection: 3D versus 2D modality,” Proceedings of the Third International Workshop in CVPR on Human Communicative Behavior Analysis (CVPR4HB 2010), San Francisco, USA.
[30]Kanade, T., Cohn, J. F., & Tian, Y. (2000), “Comprehensive database for facial expression analysis,” Proceedings of the Fourth IEEE International Conference on Automatic Face and Gesture Recognition (FG'00), Grenoble, France, 46-53.
[31]Z. Riaz, C. Mayer, M. Wimmer, M. Beetz, B. Radig, “A Model Based Approach for Expressions Invariant Face Recognition,” Proceedings of the Third International Conference, ICB 2009, Alghero, Italy, June, 2009, pp:289-298.
[32]Lucey, P., Cohn, J. F., Kanade, T., Saragih, J., Ambadar, Z., & Matthews, I., “The Extended Cohn-Kanade Dataset (CK+): A complete expression dataset for action unit and emotion-specified expression,” Proceedings of the Third International Workshop in CVPR on Human Communicative Behavior Analysis (CVPR4HB 2010), San Francisco, USA, 94-101.
[33]P. N. Bellhumer et al, “Eigenfaces vs Fisherfaces: Recognition using class specific linear projection”, IEEE PAMI, special issue on Face Recognition, 17(7), 711-720, 1997.
[34]C.-L. Lu, L.–W. Tsai, Y.-K. Wang, K.-C. Fan, “Robust Face Recognition under Illumination and Expression Variations,” Proceedings of the Ninth International Conference on Machine Learning and Cybernetics, Qingdao, July 2010.

[35]Yung-Mao Lu, Bin-Yih Liao, Jeng-Shyang Pan, “Face Recognition Algorithm Decreasing the Effect of Illumination,” Int. Conference on Intelligent Information Hiding and Multimedia Signal Processing, 2008.

[36]T. Chen, W. Yin, X.S. Zhou, D. Comaniciu, T.S. Huang, “Total Variation Models For Variable Lighting Face Recognition”, IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 28,pp:1519 - 1524, 2006.

[37]Wankou Yang, Jianguo Wang, MingwuRen, Jingyu Yang, “Feature extraction based on Laplacian Bidirectional Maximum Margin Criterion”, Pattern Recognition, In Press, Corrected Proof, 2009.

[38]M. J. Lyons, et al, “Coding Facial Expressions with Gabor Wavelets,” Proc. Third IEEE Conf. Automatic Face and Gesture Recognition, pp: 200-205, April 1998, Nara Japan.

[39]F. Roli and J. Kittler, "Multiple Classifier Systems, Third International Workshop, MCS 2002, Cagliari, Italy, June 24-26, 2002, Proceedings," in Lecture Notes in Computer Science, vol. 2364, Lecture Notes in Computer Science: Springer Ve11rlag, 2002.

[40]A. Y. Yang, A. Ganesh, Z. H. Zhou, S. S. Sastry, and Y. Ma. Fast l1-minimization algorithms and application in robust face recognition, UC Berkeley, Tech. Rep.

[41]S. J. Wright, R. D. Nowak, M. A. T. Figueiredo. Sparse reconstruction by separable approximation. In ICASSP, 2008.

[42]A. Beck and M. Teboulle. A fast iterative shrinkage-thresholding algorithm for linear inverse problems. SIAM. J. Imaging Science, 2(1):183-202, 2009.

[43]J. Yang and Y. Zhang. Alternating direction algorithms for l1-problems in compressive sensing. (preprint) arXic:0912.1185, 2009.

[44]D. Malioutove, M. Cetin, and A. Willsky. Homotopy continuation for sparse signal representation. In ICASSP, 2005.

[45] S. J. Kim, K. Koh, M. Lustig, S. Boyd, and D. Gorinevsky. Ainterior-point method for large-scale l1-regularized least squares. IEEE Journal on Selected Topics in Signal Processing, 1(4):606–617, 2007.

[46]R. Rigamonti, M. Brown and V. Lepetit. Are Sparse Representations Really Relevant for Image Classification? In CVPR 2011.

[47]Q. Shi, A. Eriksson, A. Hengel, C. Shen. Is face recognition really a compressive sensing problem? In CVPR 2011.

[48]S. Chambon, A. Crouzil. " Dense matching using correlation: new measures that are robust near occlusions". In British Machine Vision Conference - BMVC 2003, volume 1, pages 143-152, Norwich, UK, September 2003.