# MATHEMATICALLY INSPIRED APPROACHES TO FACE RECOGNITION IN UNCONTROLLED CONDITIONS - SUPER RESOLUTION AND COMPRESSIVE SENSING

BY

**NADIA AL-HASSAN**

Applied Computing Department

The University of Buckingham / United Kingdom

A Thesis

Submitted for the Degree of Doctor of Philosophy in Mathematical Science to the school of Science and Medicine in the University of Buckingham

September 2014

# ABSTRACT

Face recognition systems under uncontrolled conditions using surveillance cameras is becoming essential for establishing the identity of a person at a distance from the camera and providing safety and security against terrorist, attack, robbery and crime. Therefore, the performance of face recognition in low-resolution degraded images with low quality against images with high quality/and of good resolution/size is considered the most challenging tasks and constitutes focus of this thesis. The work in this thesis is designed to further investigate these issues and the following being our main aim:

"*To investigate face identification from a distance and under uncontrolled conditions by primarily addressing the problem of low-resolution images using existing/modified mathematically inspired super resolution schemes that are based on the emerging new paradigm of compressive sensing and non-adaptive dictionaries based super resolution."*

We shall firstly investigate and develop the compressive sensing (CS) based sparse representation of a sample image to reconstruct a high-resolution image for face recognition, by taking different approaches to constructing CS-compliant dictionaries such as Gaussian Random Matrix and Toeplitz Circular Random Matrix. In particular, our focus is on constructing CS non-adaptive dictionaries (independent of face image information), which contrasts with existing image-learnt dictionaries, but satisfies some form of the Restricted Isometry Property (RIP) which is sufficient to comply with the CS theorem regarding the recovery of sparsely represented images. We shall demonstrate that the CS dictionary techniques for resolution enhancement tasks are able to develop scalable face recognition schemes under uncontrolled conditions and at a distance. Secondly, we shall clarify the comparisons of the strength of sufficient CS property for the various types of dictionaries and demonstrate that the image-learnt dictionary far from satisfies the RIP for compressive sensing. Thirdly, we propose dictionaries based on the high frequency coefficients of the training set and investigate the impact of using dictionaries on the space of feature vectors of the low-resolution image for face recognition when applied to the wavelet domain. Finally, we test the performance of the developed schemes on CCTV images with unknown model of degradation, and show that these schemes significantly outperform existing techniques developed for such a challenging task. However, the performance is still not comparable to what could be achieved in controlled environment, and hence we shall identify remaining challenges to be investigated in the future.

*I dedicate my thesis to my Father's soul*

*And*

*To my family*

# ACKNOWLEDGMENTS

**ALLAH THE MOST GRACIOUS AND MERCIFUL:** Who gave me the energy, health and courage to spend determination and time until I completed this work.

**MY FAMILY:** My heartiest and warm thanks go to my family, for their support, patience and understanding throughout the duration of my PhD time. I begin with my soul mate, my husband ***QAHTAN*** who has been there for me every step of the way. My ***Mother*** who has not stopped praying for this work to be completed and I would like to dedicate this work and all my success to my ***Father***, who passed away before I start this work. I end with my lovely children, ***EZZULDDIN***, ***GHUFRAN*** and ***AYAT***, who have been my continuous source of hope and determination to continue, despite the difficult times I have encountered.

**MY SUPERVISORS:** I would like to express my sincerest gratitude towards my Supervisor, Professor *Sabah Jassim* for his support, patience, valuable advice, suggestions, convincing arguments, and more during the life of this thesis; I wish him all the best for the future. I would also like to thank my Supervisor Dr. *Harin Sellahewa* for his valuable comments, and useful discussions from the beginning until the end. I also owe a great debt to Dr. *Nasser AL-Jawad* for his assistance and encouragement.

**STAFF AND COLLEAGUES:** I would like to thank all the students and the staff in the Department of Applied Computing in the University of Buckingham for discussions. Special thanks go to Dr. *Hisham Al-Assam*, lecturer in the department for his advice, and help. In addition, all the love and appreciation goes to Dr. *Maysson Al-haj Ibrahim* for her advice.

I would like to thank all my friends everywhere who continually encourage and support me. Huge thanks and appreciate are also go to Dr. *Muna Mansour*, lecturer at the University of Baghdad for her help and encouragement.

**LAST AND NOT LEAST TO MY SPONSOR:** I would like to express my sincere appreciation and gratitude to the Ministry of Higher Education and Scientific Research in Iraq, to my University in Baghdad and to the Iraqi Culture Attaché in London for sponsoring my PhD program of study.

# ABBREVIATIONS

| | |
|---|---|
| AHE | Adaptive Histogram Equalization |
| AHistogram | Actual Histogram |
| BC | Bi-cubic Technique |
| CB | City Block (refers to distance function) |
| CCTV | Closed Circuit Television |
| CS | Compressive Sensing |
| CT | Computed Tomography |
| DCT | Discrete Cosine Transform |
| DFT | Discrete Fourier Transform |
| DSR | Discriminative Super Resolution |
| DWT | Discrete Wavelet Transform |
| EER | Equal Error Rate |
| ENF | Equal Norm Frame |
| FAR | False Acceptance Rate |
| FDA | Fisher Discriminant Analysis |
| FFT | Fast Fourier Transform |
| FRR | False Rejection Rate |
| FT | Fourier Transform |
| GRM | Gaussian Random Matrix |
| HD | High Definition |
| HE | Histogram Equalization |
| HH | High-High (refers to a wavelet subband) |
| HI | Histogram Intersection |
| HL | High-Low (refers to a wavelet subband) |
| HR | High Resolution |
| IBP | Iterative Back Projection |
| IHT | Iterative Hard Thresholding |
| IISR | Iterative Interpolation Super Resolution |
| IQ | Image Quality |
| K-SVD | K-Means Singular Value Decomposition |

| | |
|---|---|
| LBP | Local Binary Pattern |
| LD | Learning Dictionary |
| LD-Sp | Learning Dictionary in the Spatial Domain |
| LDA | Linear Discriminant Analysis |
| LD-WD | Learning Dictionary in the Wavelet Domain |
| LH | Low-High (refers to a wavelet subband) |
| LID | Linear Independence Dictionary |
| LL | Low-Low (refers to a wavelet subband) |
| LSM | Least Square Method |
| LR | Low Resolution |
| MRI | Magnetic Resonance Imaging |
| MSE | Mean Squared Error |
| NSP | Null Space Property |
| NP-hard | Non-deterministic Polynomial-time hard |
| OMP | Orthogonal Matching Pursuit |
| PCA | Principal Component Analysis |
| PF | Parseval Frame |
| POCS | Projection onto the Convex Sets |
| PSNR | Peak Signal to Noise Ratio |
| RIC | Restricted Isometry Constant |
| RIP | Restricted Isometry Property |
| RM | Rotation Matrix |
| RMSE | Root Mean Squared Error |
| ROM | Random Orthonormal Matrix |
| SCface | Surveillance Cameras Face |
| SD | Standard Definition |
| SGRM | Standard Gaussian Random Matrix |
| S-N | Shannon-Nyquist |
| Sp | Spatial Domain |
| SR | Super Resolution |
| SRLD | Super Resolution based on Learning Dictionary |
| SRWD | Super Resolution in the Wavelet Domain |
| STRIP | Statistical Restricted Isometry Property |
| SVD | Singular Value Decomposition |

| | |
|---|---|
| STD | Standard Deviation |
| TCRM | Toeplitz-Circular Random Matrix |
| TF | Tight Frame |
| THistogram | Theoretical Histogram |
| UIQI | Universal Image Quality Index |
| UN | Unit Norm |
| WT | Wavelet Transform |
| ZN | Z-Score Normalization |

# TABLE OF CONTENTS

# LIST OF FIGURES

# LIST OF TABLES

## List of Tables

# *DECLARATION*

This work has not previously been submitted for a degree or diploma in the University of Buckingham or any other university. To the best of my knowledge and belief, the thesis contains no material previously published or written by another person except where due reference is made in the thesis itself.

***Nadia Khazaal AL-Hassan***

# Chapter 1

## INTRODUCTION

During the last few decades, the field of digital image processing/analysis has been growing rapidly and has several applications, such as biomedical imaging, satellite imagery, image transmission, surveillance and applications in daily life. Image-based biometrics depend on many image processing and analysis techniques for the extraction of discriminating feature vectors is another high profile area of research with growing set of applications, e.g. proof of identity for access to physical/logical facilities or entitlement to services. For all such applications, the quality of the input image and the resolution are important factors to meet the objectives of the intended processing/analysis tasks. Image quality is influenced by external factors such as lighting conditions and environmental conditions that may lead to loss of contrast, addition of noise and blurriness. Image quality is also influenced by internal factors such as camera quality. Image resolution is related to the amount of detail available in the image area and is measured by the number of sampled pixels in a unit of area. In some applications, and for purposes of efficiency, the quality and resolution of a region of interest may be of more importance and concern than the entire image. For example, in face biometric applications, the region of interest is the face, which should be of good contrast and represented at a sufficiently high-resolution. In many imaging systems, under-sampling leads to degradation of image quality, as well as there being different degradations that need to be taken into account, covering for instance noise and blurring. Image blurring can be caused by several external effects, such as relative motion between the camera and the original scene, an optical system that is out of focus, camera lens, atmospheric turbulence, etc.

This thesis is primarily concerned with the investigation of mathematically inspired techniques and procedures that can overcome variation in image quality and resolution resulting from different recording conditions. We use face biometric recognition in uncontrolled recording as the closely relevant challenging application area, which could benefit from the techniques developed in this thesis. In particular, we focus on developing and testing the performance of innovative techniques of ***super resolution*** techniques to overcome resolution

limitation and to construct high-resolution (HR) images for face recognition at a distance, that lead to improved face recognition accuracy in uncontrolled conditions.

The aim of this introductory chapter is to describe the background materials and the challenges in face recognition, where these challenges are the key to understanding the motivation, to explain in general the concept of super resolution (SR) and to introduce the recent field of compressive sensing for SR adopted in our investigations. We finish this chapter by giving an overview of this research project thesis, which includes the motivation for the thesis, a statement of the main objectives of the investigations conducted for this thesis, and an outline of the approach taken to achieve these objectives. Finally, we highlight the main research contributions and outline the remainder of the thesis.

## 1.1 Biometric Systems

Over the last few decades, advances in digital technology and the availability of cost effective cameras and other sensors as well as the surge in identity theft have led to a rapid growth of interest in Biometric-based authentication. Biometric recognition systems have the objective of automatically identifying the person's identity present in the input images using human biometric identifiers such as face, iris, fingerprint and others. Compared to traditional authentication techniques which are based on either "something you know" such as password and PIN or "something you have" such as passport and ID card, biometric authentication is based on "who you are". Biometric traits are not easily reproduced, cannot be forgotten, lost, or stolen while password and ID card can be subject to these problems and can also be shared, copied and cracked.

Biometric systems require the measurement of unique information (i.e. feature vectors) characterizing the individual being enrolled or tested and later comparing/matching these samples against a database containing several candidates. This process is called feature extraction, and is the most important component. Matching is based on a distance or a similarity function defined on the space of all feature vectors. For some biometrics, the feature extraction scheme is a mapping from the biometric traits into a high dimensional vector space, such as $R^n$ or, the binary field. Examples of distance/similarity functions include Euclidian, City-block, Hamming distances and the cosine function.

Biometric recognition systems mainly operate in two stages: the registration (enrolment) stage and recognition stage. During the registration stage, face images from a number of people are captured (normally under controlled conditions), and enrolled into the system as gallery/training data (templates). The training face images set which is used to create a biometric template normally contains images of good quality and of the highest possible resolution. In the recognition stage, the face image of a test subject (probe image) which is of reasonably good size and quality is matched to the gallery data using a one-to-one or one-to-many scheme, after the feature vectors of the person's face in the gallery and probe data are extracted. Feature extraction aims to extract a set of discriminating features from the face image that can be used later to identify the person for recognition. The biometric recognition functions can be performed in two modes depending on the application context.

***Identification***: The purpose of face identification is to determine the identity of the person in the probe images based on the biometric template(s) in a database.

***Verification***: In this mode of recognition, the person declares his identity, and the biometric system compares the probe image with only one image from the template/ gallery. If the person's identity claim is correct, then he is a genuine client, while if the person claim to be someone else, he is an imposter.

An ideal biometric system should have perfect accuracy, but in practice, a biometric system can make types of errors. These types of error exist in the verification systems; these being False Rejection Rate (FRR), False Match Rate that is also known as False Acceptance Rate (FAR) and Equal Error Rate (EER). FRR occurs when two samples of the same biometric trait of an individual are not recognized as a match. FAR occurs when two samples from different individuals are incorrectly recognized as a match. While, the error EER is a connection between FRR & FAR and is computed as the point of intersection in the matching score distribution where FRR=FAR. Sometimes, the matching score distribution is not continuous and there is no cross point.

In practice, biometric error can occur due the differences in the digital feature vector representation of the biometric trait between the enrolment, where the images of HR, good size, and the recognition stages when the probe image of low-resolution (LR) small size.

Face is the most popular and natural biometric trait for recognition due to the following properties:-

1. *Universality*: face is one of the most common biometric traits possessed by all humans.
2. *Collectability*: facial information is very easily collectable.
3. *Acceptability*: face is widely considered as one of the most non-intrusive biometric features to acquire.

In fact, several other advantages exist which make the face the most preferred biometric trait such as:

- The face biometric is easy to capture even at a long distance without the person's attention.
- The face conveys not only the identity, but also the internal feelings of the person (e.g. happiness or sadness) and the person's age.

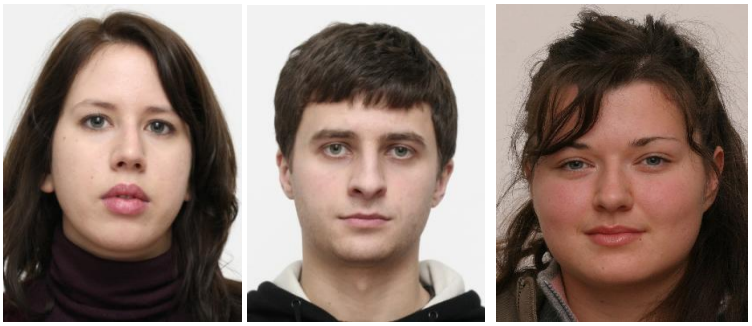
Automatic face recognition technology is used in an increasing number of applications, such as video surveillance, secure access, human/computer interface and identity management (e.g. passports, driving license cards). Face recognition systems under uncontrolled conditions by surveillance cameras are becoming essential to the establishment of the identity of a person who is at a distance from the camera and providing safety and security against terrorist, attacks, robbery, crime, etc. In the past twenty years, many different schemes for face recognition have been developed and deployed mainly for access control. They mostly differ in their feature extraction schemes, such as Principal component analysis (PCA) (Turk & Pentland, 1991), Local binary pattern (LBP) (Zhang, et al., 2012) and wavelet based face recognition (Sellahewa & Jassim, 2010).

Although face recognition is one of the most remarkable abilities of human vision and is the most widely researched area in the face biometric domain, face recognition technology has not been deployed widely in real world applications due to the many challenges that still exist such as changes in illumination, variation in poses, expressions, LR, occlusions and ageing.

The problems that are characteristic of uncontrolled recording conditions lead to intra-class and inter-class variation with adverse impact on accuracy. Uncontrolled conditions such as changing in lighting, have adversely affected the face image where the individual images (e.g. two images of the same person) appear different, and the images from different persons tend to appear more similar and that affects the performance of face recognition between the enrolment and matching stages. In fact, there is no two images of the same person are identical even when captured under similar conditions.

Consequently, face recognition is currently perceived as the least reliable technique, especially when the images are captured at a distance under uncontrolled conditions by surveillance cameras (Park, 2009), where some cameras (i.e. cameras with different resolution) often produce low quality images that make recognition more difficult and less reliable. Therefore, amongst all these challenges, the performance of face recognition in LR images is considered the most challenging task and constitutes the focus of this thesis. To illustrate this challenge, below are displayed images of three persons (captured on CCTVs) at different quality and resolution levels that are meant to be captured post enrolment together with their reasonably good quality and HR gallery images captured at the enrolment stage. Note that the three LR images exhibit different level of degradation.

Low-Resolution Probe Images



High-Resolution Gallery Images

## 1.2 Image Resolution Manipulation Techniques

A decrease in image resolution typically results in loss of facial details, which is expected to lead to a decrease in recognition rates. The uses of high quality/definition cameras (Al-Obaydy & Sellahewa, 2011), (Choi, et al., 2010) have been suggested in order to solve the problem of face recognition at a distance and improve the quality of the captured images. However, limited success and cost implications necessitate the search for software solutions.

In this section, we shall briefly discuss the concept of image resolution and describe traditional approaches to image resolution enhancement schemes.

The essence of the problem of low image resolution is not different from many experimental computations in science where the number of data samples is limited and sparse within the domain of the independent variable(s). In mathematics, such problems dealt with using optimisation and approximation techniques to obtain more samples and gain a more reliable estimation of the behaviour of the dependent variable(s). For example, when solving boundary value problems associated with partial differential equations. We normally have the solution known at a sparse set of points on the boundary of the space of interest, and more points are added to the space by a process of refinement using successive triangulation (more generally, finite elements methods). Then the solution will be expanded to these new domain points by various optimisation and approximations techniques, which depend on our knowledge of the mathematical model of the problem. Following this analogy, when one is presented with low-quality, LR images, an intuitive approach for improving the quality of the images would be to attempt to increase image ***resolution*** by adding new pixels (in between existing pixels) and using mathematical techniques such as interpolation to assign gray values of the new pixels depending on the surrounding known pixel values. Most image software and tools provide image-resizing functions, which employ bilinear and bi-cubic interpolation with some success. However, interpolation is not sufficient to improve face recognition when the captured images are not only of LR but are also of low degraded quality. A close look at the following sets of LR small images displayed on the top row of images in Figure 1.1 below are subjected to different levels of degradation and their enlarged versions obtained by the existing bi-cubic interpolation scheme and displayed on the second row. Illustrate that the quality of resulting HR images are only satisfactory when the original small images are mildly degraded. The PSNR values obtained for the enlarged images, with reference to the original high-resolution good quality image are decreasing in the range [33.72, 19.48]. The most challenging uncontrolled scenarios for face recognition involve images that are of LR and low quality that cannot be dealt with using interpolation. Modelling image degradation that covers a wide range of image quality is a difficult task that adds to the complication of face recognition.


**Figure 1.1:** Illustration of enlarged images from low-resolution images with increasing level of degradation.

The term "super resolution" refers to a process of obtaining higher resolution images from lower resolution ones. The search for higher-resolution images seems to be increasing and the ultimate aim of image SR is to produce a high-quality image that is visually clearer, sharper and contains more detail than the individual input LR images. The super resolved image should demonstrate an improvement in the perceived detail content compared with that of degraded LR images. Figure 1.2 shows sample of HR images that can be obtained by super resolving LR images subjected to different degradation.

**Figure 1.2:** Example of super resolution images from low-resolution images with increasing level of degradation.

Therefore, in order to improve the performance of face identification schemes and to deal with face recognition for LR, low quality images. SR techniques have been investigated/and developed to overcome the limitation in terms of image quality by enhancing the spatial resolution to generate HR images of the original scene, without the need for any hardware en-

hancements ( (Bannore, 2009), (Chaudhuri, 2001), and (Sroubek, et al., 2011)). This is why most of the research into image resolution enhancement has been directed towards developing image-reconstructing techniques that deliver the highest possible quality image that can be considered as the image of the same scene captured by the observed degraded image. Image reconstruction techniques have proved useful in cases where greater clarity in images is required. Some of the applications are in medical imaging (such as Computed Tomography (CT) and Magnetic Resonance Imaging (MRI)), surveillance systems with CCTV, satellite-imaging applications (such as remote sensing), and SR is required as a pre-processing for recognition.

This thesis is mainly concerned with the development of mathematically inspired SR schemes for face recognition. We shall investigate feature resolution techniques and attempt to exploit recent advances in the mathematics of compressive sensing (CS) technology. Although we focus on face recognition, the use of the developed compressive sensing SR techniques is by no means limited to face biometrics. The corresponding HR features vectors help to achieve improved recognition rates when compared to a pixel-domain SR approach. It has recently been used to improve the performance of iris recognition when dealing with LR recordings. Nguyen (Nguyen, et al., 2011) has investigated a feature domain based SR approach for iris recognition, whereby multiple LR iris images are input into the PCA (or Eigeniris) domain.

Super resolution methods can be classified into two groups:

(a) Methods that will be using a single LR image, which interpolates the pixel information available in the image.
(b) Methods that will use multiple images, and output a higher resolution image by fusing the pixel information from multiple observations of the same scene.

Mathematically, the SR approach to matching HR images in the gallery set to one or more observed LR input images $\boldsymbol{y_i}$ , $\boldsymbol{i = 1, 2, \dots, n}$ is modelled as a solution $\boldsymbol{x}$ of the matrix equation

$$\boldsymbol{y_i = SBx} \tag{1-1}$$

where $\boldsymbol{S}$ and $\boldsymbol{B}$ are the down sampling and blurring functions respectively, as illustrated below in Figure 1.3. In this model, image noise often modelled as additive noise is neglected.

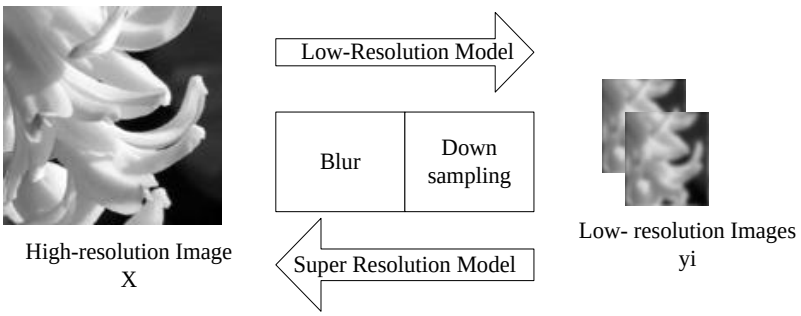



**Figure 1.3:** Super resolution Imaging Model

The SR is considered as an optimization problem that solves the above equation using a technique that attempts to reconstruct a HR image by iteratively fusing the partial information contained within a number of under-sample LR images of the pictured scene. Therefore, SR reconstruction involves up sampling of under-sampled images, thereby filtering out distortions such as noise and blur. In comparison with various image enhancement techniques, SR techniques not only improve the quality of under sampled LR images by increasing their spatial resolution but also attempt to filter out distortions. As shown in Figure 1.3, the LR images are independent images generated by blurring and down sampling the original image. Moreover, for SR to be possible each LR image contains slightly different information where a single image is unable to provide all of the information required for reconstruction. Such techniques are known to be ill conditioned, and are usually dealt with using regularisation techniques such as Lagrange multipliers (Bannore, 2009).

The above iterative optimization scheme and other SR methods along with their use of classifications are provided in the following chapters. However, in the next section we introduce the recently emerging paradigm of CS approach to sparse signal representation, which provides a new potential for SR, which is the main contribution of this thesis. As hinted above, SR, image reconstruction is an ill-posed problem due to the insufficient number of LR images that can be assumed to have been obtained by downsampling. This means that the HR image, which corresponds to the observed LR image(s), is at best sparse. Hence, recovering the sparse HR image can benefit from using CS recovery techniques. Indeed, CS has been investigated as a new SR method to recover high quality, SR images from a single LR degraded image ( (Yang, et al., 2008), (Wang, et al., 2012) and (Yang, et al., 2010)).

## 1.3 Compressive Sensing

Compressive sensing (CS) is a new theory of sampling in many applications (such as digital image and video camera), and it greatly eases the stringent limitation imposed by the traditional Shannon sampling theory on the number of samples needed to recover "small" or sparse signals. While Shannon theorem stipulates that perfect recovery of a signal require the number of samples to be at least as large as twice the highest frequency in the signal, most signal/image applications signals are compressed and represented by sparse signals using transformations that remove significant correlations and redundancies in the capture signals. Compressive sensing is an attempt to recover the sparse (compressed) signal directly by making (i.e. sensing) a relatively small number of linear measurements of the signal/image from which the significant (i.e. the nonzero) sample/pixel values of the compressed signal/image can be recovered. The relatively small set of measurements must satisfy certain properties to guarantee the unique recovery of the sparse signal. Consequently, CS is a tool for solving a class of inverse/underdetermined problems in computer vision and image processing.

The underdetermined problems can be defined as a solution of a matrix equation of the form:

$$\boldsymbol{y_{m\times 1}} = \boldsymbol{D_{m\times n}} \boldsymbol{x_{n\times 1}},\ \boldsymbol{m < n} \tag{1-2}$$

Where $\boldsymbol{y}$ is the observed signal of length m and $\boldsymbol{x}$ is the sought solution of length n, and $\boldsymbol{D}$ is the matrix of size $m \times n$ whose rows are the measurements.

This equation has no unique solution, since the number of variables in the matrix $D$ is larger than the number of rows (equations). However, according to the compressive sensing theory, there is a unique sparse solution for the equation (1-2), which can be reconstructed by a number of practical algorithms. The basic principle of sparse solution (or sparse representation) assumes (Kulkarni, et al., 2011), (Zhang, et al., 2012) that any compressible/sparse signal $x \epsilon \mathbb{R}^n$ can be represented in the form of:

$$\boldsymbol{x} = \sum_{\boldsymbol{j=1}}^{\boldsymbol{n}} \boldsymbol{D\alpha} \quad \text{Or} \quad \boldsymbol{x = D\alpha} \tag{1-3}$$

Where $\boldsymbol{\alpha}$ is a column vector of length $n \times 1$, the sparse signal $\boldsymbol{x}$ can be shown as a linear combination of only $k$ basis vectors from an overcomplete dictionary $D$ which satisfies the property of CS theorem. While most of the linear coefficients are zeros, the exceptional $k$ ones can be non-zero. The property of CS measures the degree to which each sub-matrix consisting $k$ column vectors of $D$ is close to being an isometry.

During the past few years, sparse recovery has been investigated to overcome the resolution problem of a single LR image (Yang, et al., 2010), and (Wang, et al., 2012). Therefore, it is necessary to formulate the SR in a CS framework. The LR image $y$ as illustrated above in Figure 1.3 can be written in the form of equation (1-1). i.e.:

$$\boldsymbol{y = SBx} \tag{1-4}$$

If $\boldsymbol{L = SB}$ then we can reformulate this equation as:

$$\boldsymbol{y = Lx = LD\alpha} \quad \text{Since } (\boldsymbol{x = D\alpha}) \tag{1-5}$$

For practical reasons, the application of this equation is localised to image blocks (patches) of small size to avoid having every pixel being influence by all-the image. To prevent blocky effects, it is customary to use overlapping blocks.

In SR based on CS, $\boldsymbol{x}$ is the unknown HR image patch while $\boldsymbol{y}$ is the observed LR image patch. The sparse representation $\boldsymbol{\alpha}$ of the image $\boldsymbol{y}$ in terms of dictionary $\boldsymbol{LD}$ which is denoted by $D_L$ is used to recover the corresponding HR patch $x$ in terms of dictionary $D$ denoted by $D_H$.

The main challenge is to have the dictionary $D$ satisfying appropriate CS conditions so that the sparse linear representation of a HR image patch $x$ in terms of $D$ can be recovered perfectly from the LR image patch. Yang in (Yang, et al., 2008), (Yang, et al., 2010) recently proposed a method to recover a HR image patch by using a pair of learning dictionaries $D_H$ & $D_L$ that were dependent on images and whose columns are constructed from a number of randomly selected patches of high and low resolution image sets. Recent researches in the field of CS have focused on (1) developing various approaches to designing dictionaries that satisfy the CS requirements and (2) on designing efficient algorithms to recover the sparse solution uniquely. A major part of this thesis is concerned with the study of various known CS dictionaries, and tests their performance in terms of the quality of super-resolved images. We shall also investigate the consequences for face recognition under uncontrolled conditions and test the need for image-trained dictionaries.

## 1.4 Image Quality Measures

In assessing the improvement achieved by any of the methods investigated and implemented, we need to use known image quality measures. Many image quality measures can be used to assess the degree of improvement or degradation in one or more of the attributes of images. The most widely known quality measures in the spatial domain are Peak Signal to Noise Ratio (PSNR), Mean Squared Error (MSE)/or Root Mean Squared Error (RMSE) and Universal Image Quality Index (UIQI). Many of the resolution enhancement approaches have attempted to address the issue of image resolution and are expected to maximize the quality measure PSNR or minimize the RMSE between the original hypothetical HR image and the reconstructed super resolved image. The PSNR quality measure can be defined as

$$\boldsymbol{PSNR = 20 * log_{10}(\frac{255}{\sqrt{MSE}})} \tag{1-6}$$

where the Mean Squared Error (MSE) can be computed by averaging the square intensity differences of the reference image $x$ and the reconstructed image $y$ as:

$$\boldsymbol{MSE = \frac{1}{mn}\sum_{i=0}^{m-1}\sum_{j=0}^{n-1}[x(i,j) - y(i,j)]^2} \tag{1-7}$$

On the other hand, a Universal Image Quality Index proposed by Wang and Bovik (Wang & Bovik, 2002) provides a good example of spatial domain measures that calculate the distortion/or enhancement between the recovered image and the reference image (original image) by modelling distortion as a combination of three main components: loss of correlation, illumination distortion and contrast distortion. However, in our implementation, we only need the contrast and correlation distortion components, due to the nature of the face database being used. This quality measure can be computed as:

$$\boldsymbol{UIQI = \frac{\sigma_{xy}}{\sigma_x\sigma_y} \cdot \frac{2\bar{x}\,\bar{y}}{(\bar{x})^2+(\bar{y})^2} \cdot \frac{2\sigma_x\sigma_y}{\sigma_x^2+\sigma_y^2}} \tag{1-8}$$

Where $\bar{\boldsymbol{x}}$ and $\bar{\boldsymbol{y}}$ are the mean of images $\boldsymbol{x}$ and $\boldsymbol{y}$ respectively,

$\boldsymbol{\sigma_x^2 = \frac{1}{N-1}\sum_{i=1}^{N}(x_i - \bar{x})^2}$, $\boldsymbol{\sigma_y^2 = \frac{1}{N-1}\sum_{i=1}^{N}(y_i - \bar{y})^2}$ and $\boldsymbol{\sigma_{xy} = \frac{1}{N-1}\sum_{i=1}^{N}(x_i - \bar{x})\,(y_i - \bar{y})}$.

In addition, the first component in UIQI measures the degree of correlation between two images; the second component measures the proximity of the luminance between two images and the third component measures the level of proximity of the contrast of two images. In

practice, the contrast and correlation measures of an image with respect to reference images is calculated for each window measuring $8 \times 8$ pixels in the two images, and the average of these entire blocks defines the measure quality of the entire image.

The problem with the above measures is the need to have a reference image of the same imaged object. A no-reference image quality measure is very essential when a reference cannot be found or it is impractical. The simplest and rather primitive measure that is of some use in determining the blurring level in an image is defined as follows. The square differences between an input image and further blurred version of the input image can be used as quality measure for estimating the blurring level in any input image where convolution the input image with any degradation filter such as median filter used as a reference image.

## 1.5 Motivation and Objectives

Feature Extraction is the most important step of any image-based pattern recognition system and it is meant to output a digital representation model of the imaged object, which is referred to as the feature vector. The performance of the corresponding recognition depends very much, on how well the mathematical extracted digital model reflects the physical object. Many factors influence the feature extraction process with possible adverse impact on the quality of the extracted digital feature vector which in turn could cause inter-class and/or intra-class variations. For face feature vectors, such factors include extreme variation in recording condition with respect to illumination, facial expression, pose, and aging. Over the last few decades research has been conducted to mitigate the effect of a few of these factors with varying levels of success, either by developing feature vectors that are invariant to such recording conditions or by applying some mathematical transformations to normalise the captured images and/or the extracted features. Other factors influencing the quality of face feature vectors include variation in the quality of capturing devices, distances between camera and subject as well as variation in time and environmental conditions of recording. The later set of factors are very much associated with unattended surveillance scenarios, often resulting in degraded face images that suffer from a combined effect of more than one of these factors. These factors may result in a combination of LR, small size degraded, and badly illuminated images of a non-frontal face with occlusion. Research into mitigating such variations have shown to be a very tough challenge and forms the initial main motivation for this thesis to investigate the development and use of efficient SR schemes.

Face recognition from unattended surveillance cameras is crucial to many applications such as in forensics for identifying persons suspected of terrorist and criminal acts, which may or may not be on a watch list. Whenever an accident happens or an attack occurs, there would be several videos captured by surveillance cameras placed in various locations in the surrounding and nearby areas that are usually examined manually to identify suspects and/or eliminate innocent by-passers. The problem of low image quality, as a result of surveillance cameras due to the distance between face and the camera, subject on the move, and unstable sensors has a significant impact on the performance of face recognition schemes and this could in turn undermine the quality of forensic evidences. Besides blurring and other image degradation factors, the captured images are also of LR. Therefore, improving the performance of face recognition at a distance does require a combination of procedures that improves image resolution and quality. Recognising faces when matching LR degraded small images against face images with higher resolution and of good size can be dealt with by pre-processing procedures primarily using the so-called SR methods, which aim to reconstruct a higher resolution version of the LR image.

These observations constitute the main motivation to first review existing SR techniques with a focus on improving face recognition from images of low quality and low resolution. Since the performance of face recognition depends more on the feature vectors extracted from the image rather than the spatial domain of the image, it would be more appropriate to investigate the impact of using SR schemes on the space of feature vectors. Since wavelet subbands of decomposed face images have been used as feature vectors for face recognition with good performance in controlled and some uncontrolled situations, we shall therefore test the impact of SR on face recognition when applied to the wavelet domain rather than the spatial domain.

The recent emergence of compressive sensing approaches to super resolution motivated and guided the work in this thesis for different reasons. Firstly, it would be interesting to compare the effect of using compressive sensing based super resolution approaches to that of using the traditional iterative optimisation approach. Moreover, noting that low-resolution images can be considered as sparse vectors, then perhaps provide a strong incentive to test the use of ample data-independent random/deterministic compressive sensing dictionaries for super-resolution and investigate the impact on the performance of face recognition in uncontrolled recording conditions. Indeed, the success of such an approach results in removing the need

for training sets which rely on sufficiently large and representative amounts of face images for a sample of persons.

In summary, the main aims and objectives of this thesis are:

- ❖ Investigate the suitability and implementation of various super resolution methods for enhancement of image quality/ resolution and for face recognition.
- ❖ Study the properties of compressive sensing dictionaries, for ensuring recovery of sparse vectors with emphasis on implications for face recognition in uncontrolled conditions.
- ❖ Investigate the application of compressive sensing paradigms for improving a single image resolution and quality.
- ❖ Design super resolution dictionaries in the frequency domain, to obtain biometric feature vectors of a desirable high-resolution face that could obviate the need for costly high definition cameras.
- ❖ Study various existing mathematical models for image degradation that could fit/approximate observed complex degradations in images captured from CCTV video cameras.

## 1.6 Main Contributions

This study investigates the possibility of the above objectives by conducting many experiments on different publicly available face databases such as Extended Yale B, UBHSD video and Surveillance Cameras face (SCface). These investigations have revealed promising results in the quality and accuracy of images, when compared with the well-known scale-up techniques.

The main contributions of this thesis can be stated as follows.

- Instead of traditional methods for reconstructing the high-resolution image directly, new super resolution based approaches are developed for face recognition that involve matching low quality low-resolution images.
- Developing compressive sensing based image super-resolution approaches in terms of image based learning dictionaries for improved image quality and for face recognition at different distances, sizes and degradation levels.

- Develop novel approaches for designing compressive sensing based non-adaptive dictionaries for reconstruction are investigated, and compare their performances with image-learnt dictionaries and traditional SR approaches in terms of different levels/models of image degradation and face recognition in uncontrolled conditions

## 1.7 Thesis Outline

- **Chapter Two:** this chapter is devoted to a literature review regarding face recognition and describes the problems of face recognition under uncontrolled conditions; introduce the basic information on super resolution and the most common techniques for super resolution in spatial domain.

- **Chapter Three** presents a brief introduction to the mathematical concepts of compressive sensing, and discuses the properties of compressive sensing for recovering sparse solution.

- **Chapter Four** is aimed at studying super resolution methods based on compressive sensing dictionaries to cure low image resolution and recovering images with good quality. It presents the novel approaches to construct overcomplete dictionaries which are used for reconstructing face images, and the Restricted Isometry property comparisons between different overcomplete dictionaries is also discussed in this chapter.

- **Chapter Five** is aimed at studying compressive sensing theory to develop scalable face recognition schemes that do not require training images by removing the need for training face image information to construct dictionaries based SR to recover all-face image and image feature.

- **Chapter Six:** the effect of compressive sensing and super resolution techniques on CCTV face images with unknown model of degradation is studied in this chapter.

- **Chapter Seven:** covers the general conclusions and potential directions for future work.

# Chapter 2

## FACE RECOGNITION AND SUPER RESOLUTION – A LITERATURE REVIEW

This chapter is devoted to a review of the literature and background material for face recognition in uncontrolled scenarios, whereby the input image is of low-resolution (LR), of small size and has degraded quality. The desired procedures to be applied are expected to increase the resolution and size of the image and improve the image quality. The focus of this chapter is resolution enhancement approaches to deal with LR face images presented for identification. The reviewed SR approaches include traditional interpolation like schemes as well as recently emerging approaches, which are termed super resolution schemes. In order to understand our requirements for image super-resolution for the relevant application we shall first give a brief description of biometrics and face recognition systems, as the main area of applications that incentivise our research in this thesis. Include a review of the literature on face recognition that highlights the main approaches and challenges of face recognition in general, and in uncontrolled conditions and/or at a distance in particular. Then describe traditional resolution enhancement schemes, provided by current image processing tools, highlighting the limitations of such techniques in terms of improving image quality. The chapter then introduces the concept of super-resolution, reviews existing iterative optimisation approaches and back-propagation, to be followed by reviewing the literature on compressive sensing based super-resolutions for face recognition.

### 2.1 Face Recognition

Biometric recognition systems aim to recognize persons based on the physiological characteristics that are genetically implied such as face, fingerprint, iris, DNA and hand geometry, and behavioural qualities acquired during person's life such as gait, speech and signature. Face recognition is an important and the most challenging tasks when compared to other types of biometric-based recognition. It has a number of real-world applications such as surveillance

systems, airport security and law enforcement. The performance of face recognition systems is significantly influenced by intra-class and inter-class variations, such as LR, illumination, occlusion, image noise, background and pose. Therefore, there exists a need to build reliable systems that work under different recording conditions. Several face recognition schemes have been developed and their performances have been tested.

Face recognition schemes, like every other biometric recognition, work in two stages: enrolment stage and matching/ recognition stage. In the enrolment stage, single or multiple face images from the same person for a number of people captured and then extracts a set of biometric feature and store in the form of template, to be used later in matching stage. In the matching stage, a set of feature vectors extracted from probe images and then compared with a set of stored template using a similarity or distance function to verify the claimed identity or to identify the person. For biometric systems in which feature vectors are elements in the n-dimensional space $\mathbb{R}^n$, the Euclidean and the City block distance functions and the normalised cross correlation similarity function are the most commonly used functions for matching. The Hamming distance is the natural choice for the Iris trait and biometric schemes whose feature vectors are fixed length binary strings. These distances/similarity functions will be used in this thesis for different purposes, and therefore we now briefly define these functions. Let $x = (x_1, x_2, \dots x_n)$ and $y = (y_1, y_2, \dots, y_n)$ be any two-feature vectors of length $n$ then the distance function $d_p$ between the two vectors for each $p > 0$ can be defined as the following:

$$\boldsymbol{d_p = \sqrt[p]{\sum_{i=1}^{n} |x_i - y_i|^p}} \tag{2-1}$$

If $p = 1$ then the distance function $d_1$ is called city block distance, while if $p = 2$ then the function $d_2$ is Euclidean distance. The normalised cross correlation is the cosine function between the two vectors and defined as:

$$\boldsymbol{\cos(x, y) = \frac{x.y}{\|x\| \, \|y\|} = \frac{\sum_{i=1}^{n} x_i y_i}{\sqrt{\sum_{i=1}^{n} (x_i)^2} \sqrt{\sum_{i=1}^{n} (y_i)^2}}} \tag{2-2}$$

Biometric recognition functions can be performed in two modes depending on the application context:

***Authentication or Verification***: This is a "one-to-one" comparison between the acquired individual biometric data known as probe or query data and the stored template(s) of the individual in the database. In this mode of recognition, a person claims an identity and the bio-

metric system aims to verify the authenticity of the claimed identity; i.e. this system simply answers the question "am I the person whom I claim to be?" Therefore, verification might be used, for example, when a user wants to access their bank account or personal computer.

***Identification**:* This is a "one-to-many" comparison between the acquired biometric data and all biometric templates in a database. In this mode of recognition, a biometric system aims to identify an individual by searching the set of available identities, or to reject them if it cannot find the identity because the identity is not declared or is unknown to that system. Throughout this thesis, the focus is on the identification scenario, but all the results can be extended easily to verification.

### 2.1.1 Feature Extraction for Face recognition

Feature Extraction is the most important step of any image-based pattern recognition system and is meant to output a digital representation of the imaged object, which is referred to as the feature vector. The performance of the corresponding recognition depends very much, on how well the mathematical model of the extracted digital feature vector reflects the physical object. Variations in the recording conditions, including the ones described above, has a significant bearing on the success of the corresponding face recognition scheme. In this section we briefly describe the most commonly used face feature extraction schemes ending with the wavelet-based scheme that is adopted in this thesis.

A face image is typically represented by a high dimensional array of pixels, each of which is represented by an integer in the cyclic group $Z_{256}$ for a monochrome image, or a 3-dimensional vectors of such integers for RGB (or any tricolour) image. For a reasonable size monochrome, face image, the processing/analysis of its content is a computationally demanding task let alone extracting the desired face feature vectors. Therefore, it is essential to apply a dimension reduction procedure to remove redundant data without losing significant features. In the statistical pattern recognition literature, Principle Component Analysis (PCA), also known as the Eigenface method (Turk & Pentland, 1991), is the most popular technique used to reduce the large dimensionality of face images by representing the face image by a linear combination of a carefully selected reasonable size set of basis vectors. The required set of vectors are meant to encapsulate the most important information content of face image and is obtained by solving the solution of an eigenvalues problem associated with a training set of face images. In fact, the basis vectors are the eigenvectors of the covariance matrix $S_{cov}$,

which is calculated from the training set of face images (called the gallery set) in the high-dimensional image space. The covariance matrix can be defined as:

$$\boldsymbol{S_{cov}} = \frac{\mathbf{1}}{\boldsymbol{N}} \sum_{\boldsymbol{i=1}}^{\boldsymbol{N}} (\boldsymbol{z_i} - \bar{\boldsymbol{z}})\,(\boldsymbol{z_i} - \bar{\boldsymbol{z}})^{\boldsymbol{T}} \tag{2-3}$$

where $\bar{z}$ is an average face image, N number of face images and the eigenvalue problem can be solved as:

$$\boldsymbol{\Lambda} = \boldsymbol{V^T S_{cov} V} \tag{2-4}$$

where $V$ is the eigenvector matrix of $\boldsymbol{S_{cov}}$ and $\boldsymbol{\Lambda}$ is the diagonal matrix containing eigenvalues of $\boldsymbol{S_{cov}}$ on its main diagonal. The eigenvectors are then stored so that their corresponding eigenvalues are in descending order. These vectors are called Eigenface, and when a new image or vector is presented for face recognition, it is projected into the new feature space and a match is found using a nearest-neighbour classifier (or any other classifier). Despite PCA success in reducing false acceptances, on the other hand false rejections due to the within class variations caused by illumination and pose remain in the Eigenface scheme (Jassim, 2010). In addition, this method unfortunately requires the use of a sufficiently large training set of multiple face images of the enrolled persons. An alternative approach to PCA, which is based on face class (category) information for reducing the dimensionality of the feature space is the Fisher Discriminant Analysis (FDA) or as it is also known, Linear Discriminant Analysis (LDA) (Sellahewa, 2006). This method is used to maximize the ratio of the determinant of the between-class scatter to that of within-class scatter.

Another type of facial feature extraction is Local Binary Pattern (LBP). This technique originally developed for local spatial texture description and proved itself a good measure for texture classification. Based on the LBP operator, each pixel of an image is labelled with an LBP code and the histogram of each label can be used as a texture descriptor of the considered region (Maturana & Soto, 2009). The LBP code of the centre pixel in the neighbourhood is obtained by converting the binary code into a decimal one. In general, the feature vector can be extracted by dividing the facial image into several non-overlap blocks, then for each block, the LBP histograms are computed. Finally, all the LBP histograms are concatenated into a single vector. However, this algorithm cannot capture all the different types of structure because the central pixel in the LBP code is not included in the calculation of local structure. Frequency transforms are commonly used to extract frequency information from any signal as a pre-processing procedure for further processing and/or analysis. Frequency information

content conveys richer knowledge about features in images that could be exploited to complement knowledge gained from the spatial information. Fourier Transform (FT) is among the most widely used tool for analysing signals of any dimension including images. The Fast Fourier Transform (FFT) and the Discrete Cosine Transform (DCT) have been used extensively for compression and other feature extractions. While the DFT provides information about the frequency content of the decomposed signal, it does not provide localisation information. The Wavelet transforms (WTs) and in particular its discrete version (DWT) is another example of frequency domain transformations that have been used with significant success in image processing and analysis tasks, including face recognition. The DWT decomposes any signal into subbands of different frequency ranges and allows perfect reconstruction. The main advantage of using the DWT over the DFT is its ability to simultaneously provide both frequency as well as spatial support (Bovik, 2009). Moreover, DWT is very efficient in isolating redundant information. Consequently, in this thesis the DWT is our preferred choice of extraction scheme used for face recognition.

DWT is a multi-resolution analysis technique that decomposes a signal into low and high frequency subbands, each of which can be transformed repeatedly, providing multiple resolutions of the signal at different scales and different ranges of frequencies. The WT of any signal is a representation of the signal in terms of a family of orthonormal wavelet bases obtained from a single wavelet function, called a mother wavelet, through repeated translation. Wavelet Transform consists of different filters that have been designed/and used in the literature for various signals and image processing. The most prominent examples are the Daubechies (db) (Daubechies, 1990) and its family filters including db2, db4, db6, and db8 of length 2, 4, 6 and 8 respectively. Daubechies 2 filter, which is also known as Haar filter, is a piecewise constant function and can be defined as:

$$\boldsymbol{h}(\mathbf{t}) = \begin{cases} \mathbf{1} & \mathbf{0} \leq \mathbf{t} < \mathbf{0.5} \\ \mathbf{-1} & \mathbf{0.5} \leq \mathbf{t} < \mathbf{1} \\ \mathbf{0} & \mathbf{otherwise} \end{cases} \tag{2-5}$$

The Haar wavelet function decomposed the signal into approximation subband containing the low frequencies by averaging the coefficient and a detail subband containing the high frequencies by differencing the coefficients.

The separable property of the DWT makes the implementation of the 2-dimensional wavelet transform (DWT) of images equivalent to a successive implementation of the 1-dimensional

DWT in two orthogonal directions. It is customary to apply the DWT firstly in the horizontal direction across the rows of the input image into approximation and details subbands. The approximation subband, commonly referred to as the low-subband, represents low frequencies in the horizontal direction, while the details subband, called the high-subband, represents the high frequencies in the horizontal direction. This would be then followed by decomposing these two subbands in the column direction, thereby decomposing each horizontal subband into two low and high subbands. Therefore, the 2-dimentional signal will be decomposed into 4-subbands: low frequency subband ($LL$) and high frequency subbands ($LH, HL,$ & $HH$). The $LL$ subband represents the low frequencies in both horizontal and vertical directions, $LH$ and $HL$ subbands represent the high frequencies (indicating significant features such as edges) in the vertical direction and in the horizontal direction, respectively. The high frequency $HH$ subband represents the highest frequencies in the image and holds information about "diagonal features" in the image. The recursive decomposition is performed only on the approximation coefficients and Figure 2.1 is an example of level three of the pyramid decomposition scheme.

Numerous wavelet filter banks can implemented in a variety of decomposition schemes. The Haar wavelet filter has been selected for use throughout this thesis due to its simplicity. At a resolution level of $k$, the pyramid scheme decomposes an image I into $3k + 1$ subbands $(LL_k, LH_k, HL_k, HH_k, \dots \dots LL_1, LH_1, HL_1, HH_1)$ at a resolution level of $k$ where the $LL_k$ is considered as the $k$-level approximation of image I, while $LH_k, HL_k, HH_k$ captures horizontal, vertical and diagonal features of the image.

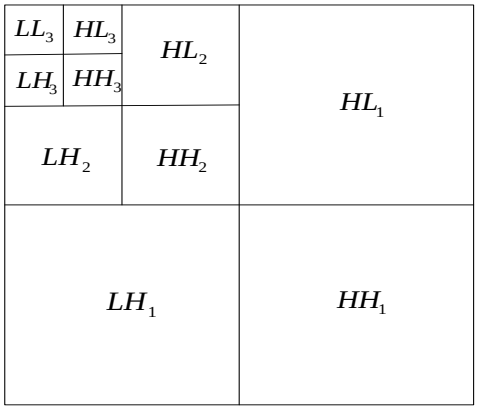



**Figure 2.1:** Wavelet Transform Level 3.

The coefficient in each subband of wavelet transformed face image (i.e. super resolved image) is used as a feature vector representation of the face for recognition purposes (Sellahewa &

Jassim, 2009), (Sellahewa & Jassim, 2010) with various accuracy rates. In the recognition stage, the feature vectors of gallery and probe images are typically compared by calculating a distance score between the two feature vectors (using City Block distance). The probe's identity is classified according to the nearest neighbour criterion. The City Block distance ($CB$), which is also called the Manhattan distance function, between two vectors $\boldsymbol{x} = \begin{bmatrix} \boldsymbol{x_1} \\ \vdots \\ \boldsymbol{x_n} \end{bmatrix}$ and $\boldsymbol{y} = \begin{bmatrix} \boldsymbol{y_1} \\ \vdots \\ \boldsymbol{y_n} \end{bmatrix}$ is given by:

$$\boldsymbol{CB(x,y) = |x_1 - y_1| + |x_2 - y_2| + \cdots + |x_n - y_n|} \tag{2-6}$$

### 2.1.2 Face recognition – Uncontrolled Scenarios

In this section the literature on face recognition in uncontrolled conditions, such as in the case of unattended surveillance situations or at a distance, is discussed and reviewed. Developing such schemes that perform well is a tough challenge (Tistarelli, et al., 2009). This is due to many factors, which might influence biometric feature extraction procedures. The performance of face recognition schemes (in which feature extraction is based on using training sets of images) is strongly dependent on how representative of the real world the training set is. i.e. how much different the images in the matching stage are from those in the training set in terms of recording conditions that influence among other things image quality, resolution and illumination. For many face recognition algorithms, the recognition rate drops dramatically when there are significant variations in the image spatial resolution and/or image quality between the training and the matching stages in terms of the influencing factors. Face recognition in uncontrolled conditions refers to applications for which one or more of the following non-exhaustive list of sources of variations in recorded conditions could cause degraded performance of face recognition schemes.

(1) *Illumination*: Variation in illumination could change the appearance of the face drastically. In the case of face images, illumination depends on the alignment of source of light with respect to the camera. Badly illuminated images suffer from appearances of shadows. The change in appearance is often more significant between differently illuminated images of the same person than between similarly illuminated images of different persons. This problem makes face comparison very difficult.

(2) *Pose*: This is dependent on the position of the camera and orientation of the face. When a camera tracks the face of a walking person and/or the camera is at an elevated position above the tracked person's head then the different image frames are associated with the different pose. The latter case is associated with images from most ordinary security cameras, thus the viewing angle of faces has a slight downward tilt. Therefore, face features such as (i.e. eyes, mouth, nose, and chin) in the captured images are affected dramatically and may not be completely visible.

(3) *Facial Expression*: Humans are capable of showing their feelings or state of health in a manner, which changes the appearance of their face in certain ways. These expressions include happiness, sadness, disgust, fear, laughing, embarrassment, etc. These expressions are dependent on many factors such as cultural background and gender, and are affected significantly by extreme variation in lighting conditions. For each expression, the face appearance change is a result of tension or relaxing certain muscles in the certain parts of the face. Humans are capable of recognising the various expressions on other people's faces. In the literature regarding automatic machine recognition two different challenges, exist in relation to this factor: person identification regardless of their facial expression and expression identification regardless of the person presented.

(4) *Occlusions*: In some unconstrained environments parts or all of the face image may be occluded (i.e. invisible) either as a result of head orientation away from camera (deliberately or not), the presence of obstacles, or as a result of wearing caps/hoods, sunglasses and scarves. This makes face identification somewhat more difficult due to missing features and data.

(5) *Low Image Resolution*: This is the case when the number of samples in the digitised spatial domain is small as a result of using cheap, low quality cameras or the face image is captured at a long distance from the camera when the view area is wide/large and the proportion of the face in the whole image is small. For example, when surveillance good quality cameras are deployed which are usually in public spaces such as metro stations and airports, faces recorded from a significant distance are of small size relative to the captured image. Therefore, the facial image is always at LR, which degenerate both the performance of face detection and the face recognition schemes. Generally, the LR images contain limited information and details are lost.

(6) *Blurred image*: In many cases of face recognition at a distance, the face image is blurred with little contrast. Blurring gets worse as the distance to the camera increases. This challenge is not unique to face recognition at a distance but happens when the person is moving fast, the camera is unstable and the optical system is out of focus, or may be due to other problems like abnormal weather and atmospheric conditions such as thermal waves and wind speed. However, these challenges are somewhat customized and more pronounced in face recognition at a distance.

Face recognition in uncontrolled condition when there is an extreme variation in a single influencing factor while the other factors are minimally varied from the neutral condition. There has been extensive amount of research and publications that deal with the challenge in such scenarios with varying level of success. In the rest of this section, we shall review the literature in relation to this type of scenario for most of the above influencing factors. Typically, the large intra-class variations due to illumination, as well as expression variations, (Sellahewa & Jassim, 2008) make face recognition a challenging problem in all face biometrics applications. The case of extreme variation in illumination from training images has been studied extensively and many approaches have been developed to mitigate these variations (e.g. see (Sellahewa, 2006), (Abboud, 2011), (Al-Assam, et al., 2012), (Al-Assam, 2013)). It has been found that these variations yield degraded images corrupted by shadows causing significant variation in face feature vectors, which in turn is manifested by loss of discriminating power. Pre-processing illumination normalization techniques, e.g., histogram equalization (HE) with the aim of transforming recorded images to create the impression of being recorded under normal illumination/conditions, have been used to address the effects of varying conditions and normalise intra-class variation. However, normalizing well-lit face images could lead to a decrease in recognition accuracy. Therefore, Sellahewa et al in (Sellahewa & Jassim, 2009) proposed the quality based adaptive HE technique to deal with the problem of variable illumination by partitioning the face image into four parts and HE used to normalize only the region where the luminance distortion is higher than a predefined threshold. This approach improved the performance of face recognition under varying lighting conditions and produced higher results than non-adaptive HE scheme. On other hand, the adaptive fusion approach to multi score of resulting wavelet based feature representations (Sellahewa & Jassim, 2010), (Abboud, 2011) focused on overcoming the problem of varying illumination conditions and to increase recognition accuracy. This approach is based on the image quality of the sample image, which has been used to select fusion parameters. In a further approach,

error correction with binaries face feature vector (Hussein, 2014) is recently proposed and used as an alternative to pre-processing normalization techniques in mitigating the effect of variations in recording conditions on automatic face recognition. Error correction means the ability to reconstruct the original data by the receiver.

In some applications where high level of accuracy are necessary, such as e-passports and border/airport checks, in this case the problem of variation illumination can be solved through recognition systems, where the system restricts the variation by imposing normal illumination i.e. imposing strict control over recording conditions, whereby the person is recorded in a well-lit booth.

Expanding the gallery to include the number of samples/templates for each individual recorded at different orientations has been used to deal with pose and expression variation. However, this approach requires huge storage space. In addition, in some application such as crime fighting it may not be possible to capture different poses or expressions.

The resolution of video frames/ images plays a vital role in face recognition from a distance where in some applications, video surveillance for example, we may only have access to low resolution and/or degraded images captured at a distance where the face is small relative to the field of view. In terms of "distance" from user to the camera, face recognition systems can be categorized into *near or close-distance, middle-distance, and far-distance*. For near-distance, the camera might capture an image of good quality and stable face images, while at far distance the images captured are usually small and of low quality. A few research works have focused on the influence of variations of image resolution and modest image degradation, (e.g. (Al-Obaydy & Sellahewa, 2011), (Bailly-Bailliére, et al., 2003), (Phillips, et al., 2009) & (Hu, et al., 2012)). The main developed approaches to deal with these challenges can be categorized essentially into two groups: (i) acquiring high-resolution face images using a special camera system and (ii) generating a super resolution face image from a given LR. High definition (HD) video cameras has been introduced recently (Al-Obaydy & Sellahewa, 2011) as a new video standard, that provides high quality video with high resolution as opposed to an LR standard definition (SD) video camera. Recent studies (Bailly-Bailliére, et al., 2003), (Phillips, et al., 2009) have shown that using HR video results in better face recognition accuracy. Al-Obaydy et al. (Al-Obaydy & Sellahewa, 2011) investigated the use of HD video cameras to recognise faces from a distance in indoor/outdoor-recorded conditions, while SD video camera has been used for comparisons. This study indicated that the high-

resolution images from HD cameras improved the recognition accuracy and produced better results than the images from SD cameras when the gallery and probe images consist of faces captured from a distance. However, at close distance HD cameras do not provide an added benefit in terms of face recognition where the face data obtained from standard/low quality image (i.e. from SD camera) provides higher results. Therefore, the choice of HD or SD camera depends on the quality of the gallery and the probe images presented for identification. Consequently, when low-quality images have been obtained, resolution enhancement methods may be a necessary step for accurate face recognition at a distance as an alternative to the costly high definition camera. However, if only a small face image with low quality (i.e. degraded image) is available for the recognition, it is questionable that traditional resolution enhancement alone can improve the performance. We shall investigate this and establish, later in this thesis, that even adding a traditional quality improvement step (known as image restoration) can only work with modest image degradation. Super-resolution is the process that deals with LR to improve the quality of the images, and has been used in such situations as a ***pre-processing*** technique to enhance either the resolution of a single LR image or the resolution of multiple LR images (Bannore, 2009), (Irani & Peleg, 1991) for face recognition at a distance. Figure 2.2, depicts a framework for face recognition that deals with an LR problem by a three-stage process: super resolution, feature extraction and face recognition system.

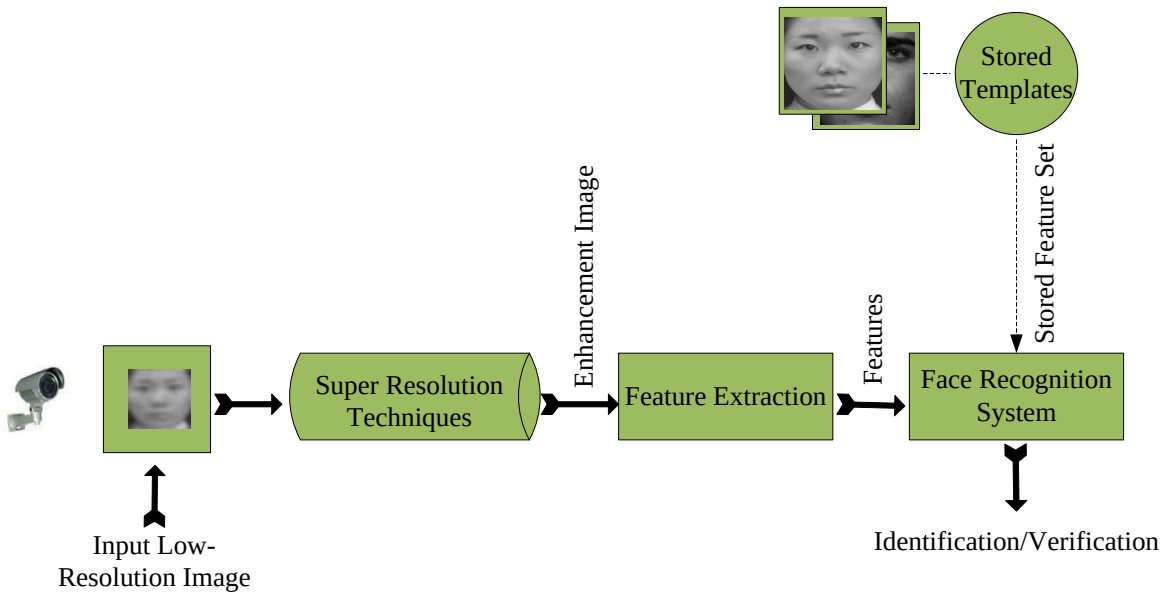



**Figure 2.2:** General configuration of a face recognition system at a distance.

In the case of unattended CCTV surveillance, the variation in distances from the camera is not the only factor to have adverse impact on the quality and resolution of the images. One of the main challenges in this respect is the difficulty in determining a mathematical model for image degradation that encompasses the combined effect of various conditions such as resolution, blurring, lighting and poses. Developing face recognition schemes that perform well in a range of uncontrolled conditions is a tough challenge (Tistarelli, et al., 2009). The use of the super-resolution approach is the obvious approach and will be investigated in this thesis. The description of super-resolution and the review of the relevant literature will be covered in Section 2.3. The remainder of this section is devoted to reviewing existing image spatial resolution enhancing and image restoration schemes.

## 2.2 Traditional Resolution Enhancing Techniques

Image resolution refers to the number of samples measured when the image is digitised. This is obviously influenced by the size of the imaged scene compared to the display size, the sampling frequency and the sensor quality. Central to understanding the image resolution problem and the implications of low image resolution, is the question about the necessary number of digital samples that guarantees the recovery of an image, which exhibits the continuity of the scenery. Surely, more samples means more details and nearer to a continuous signal. However, there are limitations to how much improvement in detail is necessary.

**Theorem 1.** *The Shannon-Nyquist (S-N) Sampling theorem*.

An under-sampled image can be perfectly reconstructed/interpolated if the sample frequency $f_s$ is greater than twice the highest frequency $f_h$ of the original image.

**Proof.** (See (Jerri, 1977))//

The S-N sampling theorem provides a sufficient, but not necessary, condition for perfect reconstruction. If $f_s < 2f_h$ then all frequency components higher than half the sampling frequency are reflected as lower frequencies in the reconstructed image, i.e. a sparse image.

Enhancing resolution means increasing the number of samples measured either by using more sophisticated high definition sensors or by developing software algorithms that insert new samples and exploit knowledge about the scenery in assigning values to the additional samples. Note that neighbouring image pixel values do not change in a random manner; otherwise we will have noisy images and difficult to recognise structures. Knowledge about image

scenery is related to the mathematical models that are used to reflect the changes in regional and neighbouring pixel values.

Resolution enhancement is required in two different types of scenarios: (1) when we have a single image of LR quality and (2) when two or more LR images of the same object are captured by two or more (CCTV) cameras from different angles. Traditionally, the first case is dealt with by interpolation whereas the process of registration deals with the second case. We shall now describe both processes used for increasing image resolution. Here the aim is not to improve image quality per se, but rather compensate for a rather low sampling rate.

### 2.2.1 Interpolation

Extrapolation and Interpolation (Levy, 2010) are subfields of numerical analysis and are important fields in mathematics. Extrapolation refers to the method of constructing new signal (data points) outside set of known data points (available signal). Interpolation, refers to the process of estimating the pixel intensities when an image is resized from a lower resolution/ or under-sample to a higher resolution, i.e. find the pixels between known data points. The image interpolation technique which is known as image resizing, image re-sampling, digital zooming and image magnification/or enhancement is an important image processing operation applied in diverse areas such as medical images, reconstruction, computer graphics, as well as being more useful for super resolution. Interpolation based super resolution has been used for a long time and it is requires to interpolate the missing pixels in the HR grid $\widehat{H}$ when the number of the LR pixels is insufficient to fill all HR pixels. In fact, image resizing procedures in most image processing tools (e.g. as Photoshop) use a variety of interpolations depending on the assumed mathematical model of regional variation in neighbouring pixel values (e.g. linear, quadratics or other polynomial functions/splines).

The interpolation process can be described in terms of $1D$ signal in the spatial domain by the following convolution formula:

$$\boldsymbol{f(x) = f_k(x) * h_1(x)}$$

where $\boldsymbol{f_k(x)}$ is the sampled signal and $\boldsymbol{h_1(x)}$ is the interpolation/reconstruction kernel which is satisfies the two properties (Bannore, 2009):-

(1) Symmetric i.e. $\boldsymbol{h_1(x) = h_1(-x)}$ and,

(2) It is equal zero for all non-zero integer numbers $x$ otherwise it is equal to one as:

$$h_1(x) = \begin{cases} 1 & if\ x = 0 \\ 0 & if\ x \neq 0 \end{cases}$$

The most commonly used interpolation techniques are the nearest neighbour, the bilinear, and the bi-cubic interpolation. These are regarded as the most attractive methods and are widely used due to their ease of computation. Figure 2.3 illustrates an image interpolated with various interpolation techniques.

(1) ***Nearest Neighbour Replacement***

The Nearest Neighbour interpolation method uses the value from the nearest pixel, i.e. replaces the interpolate point with the nearest neighbouring pixel, where the value of the new point is taken as the value of the old coordinate point which is located nearest to the new point. In other words, the variation of neighbouring pixel values is modelled as a *step function*. The nearest interpolation kernel is defined as the following:

$$h_1(s) = \begin{cases} 1 & if \quad s\ \epsilon\ [-0.5, 0.5] \\ 0 & elsewhere \end{cases} \tag{2-7}$$

Where **s** is the distance between the origin pixel and the interpolate point/pixel. This approach is very simple and easily implemented. However, it performs poorly and the resultant aliasing or blocky effect makes the image quality unacceptable for most high quality imaging applications, (see the examples below). One can see that even when the LR image is of reasonably quality this approach introduces some degradation in the form of aliasing/blocky effect.

(2) ***Bilinear Interpolation***

The bilinear interpolation is based on the assumption that the variation of neighbouring pixel values is modelled as a *step function.* Consequently, the missing pixel values can be determined as a weighted average of four neighbouring pixel values that are surrounding the unknown pixel. In other words, the one dimension linear interpolation technique applied sequentially in both directions (horizontal and vertical). The missing pixel value in each direction is evaluated in terms of its immediate neighbouring pixels (say $P_1, P_2$), in the given direction, using the interpolation function/kernel given below:

$$h_1(s) = \begin{cases} 1 - s & 0 \leq s < 1 \\ 0 & elsewhere \end{cases} \tag{2-8}$$

Suppose $\boldsymbol{x}$ is any unknown point at the sample data then:

$$\boldsymbol{x = c_1 P_1 + c_2 P_2 = \frac{P_1 + P_2}{2}}$$

Since $$\boldsymbol{c_1 \, and \, c_2 = 1 - s = 0.5 \quad if \; s = 0.5}$$

The following LR images and their bilinear interpolated images illustrate that this method performs better than the nearest neighbour method. For example, when applied to good qualities LR images the aliasing/blocky effect is less visible. However, some degradation is still visible in the form of blurring, especially when the input LR image is of low quality.

(3) ***Bi-cubic Interpolation***

This interpolation method is based on the assumption that the variation of neighbouring pixel values is modelled as a *cubic spline function.* Therefore, missing pixel values can be determined sequentially in the horizontal and then in the vertical direction by a cubic spline determined of the closest four known neighbouring pixels in that direction. The one-dimensional bi-cubic interpolation function in any direction is evaluated using the piecewise cubic polynomials (i.e. cubic spline) determined by the four neighbouring pixels ($\boldsymbol{P_1}$, $\boldsymbol{P_2}$, $\boldsymbol{P_3}$, $\boldsymbol{P_4}$). The following is the general expression for this interpolation technique (see (Bannore, 2009)):

$$\boldsymbol{h_1(s)} = \begin{cases} \frac{3}{2}\mathbf{s^3} - \frac{5}{2}\mathbf{s^2} + \mathbf{1} & \mathbf{s} \in [\mathbf{0}, \mathbf{1}) \\ -\frac{1}{2}\mathbf{s^3} + \frac{5}{2}\mathbf{s^2} - \mathbf{4s} + \mathbf{2} & \mathbf{s} \in [\mathbf{1}, \mathbf{2}) \\ \mathbf{0} & \mathbf{elsewhere} \end{cases} \tag{2-9}$$

where $\boldsymbol{s}$ is the distance between the point to be interpolated and the point being considered. The kernel $\boldsymbol{h_1(s)}$ is zero for all integer values of its argument and is one at zero. Suppose $\boldsymbol{x}$ is the point to be interpolated, then if $\boldsymbol{s = 0.5}$ then $\boldsymbol{x}$ is equal:

$$\boldsymbol{x = c_1 P_1 + c_2 P_2 + c_3 P_3 + c_4 P_4}$$

Where

$$\boldsymbol{c_1 = -\frac{1}{2}(1+s)^3 + \frac{5}{2}(1+s)^2 - 4(1+s) + 2}$$
$$\boldsymbol{c_2 = \frac{3}{2}s^3 - \frac{5}{2}s^2 + 1}$$
$$\boldsymbol{c_3 = \frac{3}{2}(1-s)^3 - \frac{5}{2}(1-s)^2 + 1}$$
$$\boldsymbol{c_4 = -\frac{1}{2}(2-s)^3 + \frac{5}{2}(2-s)^2 - 4(2-s) + 2}$$

In addition $(\mathbf{1}+\boldsymbol{s}), \boldsymbol{s}, (\mathbf{1}-\boldsymbol{s}) and\ (\mathbf{2}-\boldsymbol{s})$ are the distance between $\boldsymbol{x}$ and the points $\boldsymbol{P_1}, \boldsymbol{P_2}, \boldsymbol{P_3}\ and\ \boldsymbol{P_4}$ respectively (Parker, et al., 1983), (Keys, 1981).

The following LR images and their bi-cubic interpolated images illustrate that this method outperforms both previous methods when applied to good quality LR images, with barely any visible degradation. In fact, the output-enlarged image is slightly sharper than that produced by bilinear interpolation, while the bilinear method introduces blurring effects and does not have the aliasing/or blocky effect shown by nearest interpolation. Therefore, in all the experiments we will focus on the bi-cubic (BC) interpolation technique.

(a) (b) (c) (d)

**Figure 2.3:** Well-known interpolation methods: the first column (a) original LR images with good and bad quality, the second column (b) nearest interpolation, the third column (c) bilinear interpolation, and the four column (d) bi-cubic interpolation.

### 2.2.2 Registration

Image registration is the process of spatially aligning two images of the same object/scene so that pixels in the two images correspond to the same region of the scene. It is necessary for image applications that require a pixel-by-pixel comparison. Image registration essential for different images is obtained using different sensors or at different times. In addition, image registration is the obvious tool for fusing and/or comparing the spatial information content of different images, which requires determining a pixel-to-pixel correspondence (Sabuncu, 2004). Moreover, it is applicable to estimating scene motion for each image with reference to one particular image.

Image registration is therefore applicable to resolution enhancement, when two LR images are available. The first step in obtaining an HR image that combines these two LR images is to find the relative transformations between the input LR images. Each pixel from each LR image is placed onto an HR composite grid $\widehat{H}$ based on the registration information. A miss-registration of images will cause the pixels to be placed incorrectly in the composite grid, which causes loss of important information and has disastrous implications on overall SR performance.

### 2.2.3 Image Restoration

The quality of the interpolated SR image generated by any of the above interpolation algorithms is inherently limited by the amount of data available in the image and its quality. This is due to the fact that the frequencies captured in the interpolated image are a combination of the frequencies existing in the LR input image and therefore interpolation cannot recover the high frequency components during the sampling process. To overcome this problem we need to work on the frequency domain of the LR images using Fourier or wavelet transforms.

Image restoration or image de-blurring refers to the process of improving the information content present in the degraded or blurred image by estimating a mathematical model for the degradation. Therefore, the object of restoration is to remove identified distortion from the observed image and provide a best possible estimate to the original undistorted image. There are two well-known restoration techniques, namely inverse filter and wiener filter. The simplest approach is the inverse filter (Bovik, 2009), where the degradation function H is inverse to the blurred image G in the frequency domain, i.e. the Fourier transform estimate $\widehat{\boldsymbol{F}}(\boldsymbol{u},\boldsymbol{v})$ of the original image can be computed by

$$\widehat{\boldsymbol{F}}(\boldsymbol{u},\boldsymbol{v}) = [\frac{\mathbf{1}}{\boldsymbol{H}(\boldsymbol{u},\boldsymbol{v})}]\boldsymbol{G}(\boldsymbol{u},\boldsymbol{v}) \tag{2-10}$$

The restored image $\hat{f}$ in the spatial domain can be recovered by the inverse Fourier transform of the frequency domain estimate $\widehat{\boldsymbol{F}}(\boldsymbol{u},\boldsymbol{v})$. However, in some cases the inverse filter may not exist due to the degradation function $\boldsymbol{H}(\boldsymbol{u},\boldsymbol{v})$ containing zero or near zero values at the frequency domain coordinate $(\boldsymbol{u},\boldsymbol{v})$.

To overcome this problem in 1942 Weiner et al (Bovik, 2009) proposed a new restoration method, which is called minimum mean square error or wiener filter. Wiener restoration filter

is the most important technique for removing blur, and estimate $\widehat{\boldsymbol{F}}(\boldsymbol{u},\boldsymbol{v})$ of the transform of the original image can be computed by

$$\widehat{\boldsymbol{F}}(\boldsymbol{u},\boldsymbol{v}) = \left[\frac{\mathbf{1}}{\boldsymbol{H}(\boldsymbol{u},\boldsymbol{v})}\frac{\boldsymbol{H}^{*}(\boldsymbol{u},\boldsymbol{v})}{|\boldsymbol{H}(\boldsymbol{u},\boldsymbol{v})|^{2}+\boldsymbol{K}}\right]\boldsymbol{G}(\boldsymbol{u},\boldsymbol{v}) \quad (2\text{-}11)$$

where the function $\boldsymbol{H}^{*}(\boldsymbol{u},\boldsymbol{v})$ is the complex conjugate of the degradation function $\boldsymbol{H}(\boldsymbol{u},\boldsymbol{v})$ and $\boldsymbol{K}$ is the scalar constant.

The following images illustrate the benefits of applying these two methods on low and badly degraded images.


Figure 2.4: Restoration Techniques for images with low/severe blur degradation.

## 2.3 Super Resolution

Super Resolution is a promising digital imaging technique, which attempts to obtain an image with higher resolution from input LR image/images that are also of degraded quality. It offers more spatial domain detailed version of images captured at a distance that are usually of low-resolution, and subject to quality degradation due to a variety of factors including sensor quality and/or environmental factors at the time of recording. Consequently, SR covers the combined objectives of resolution enhancement techniques (interpolation and registration) as

well as those of image restoration techniques. In fact, conventional SR techniques incorporate those techniques to deal with the simultaneous challenges of low-resolution and low quality images. The goal of SR algorithms is to recover lost high frequency components in a single (or multiple) LR image(s) in a way that approximates the perceived/imagined HR image as closely as possible.

Super-resolution enhancement methods attempt to accumulate information from a sequence of low quality/ resolution images to reconstruct one super resolved image. However, enough images would not always be available. Therefore, during the last few years, some researchers (Freeman, et al., 2002), (Zhang, et al., 2012), (Kulkarni, et al., 2011) have focused on reconstructing HR images from a single low quality low-resolution input image.

The use of super resolution techniques is not limited to surveillance scenarios, but can be used for other purposes such as image inpainting. In (Meur & Guillemot, 2012) SR has been used as post processing to improve inpainted image quality after inpainting technique, where inpainting technique is first applied to the input image for filling the missing region.

Image inpainting is an image processing challenge that also deals both requirements of SR, i.e. filling missing data and restoring image quality that may result from the inpainting first steps. Inpainting is concerned with the restoration of damages and cracks in old images/photographs. The concept of inpainting is essentially a type of interpolation and the word "inpainting" refers to filling-in missing region (holes or damage such as cracks or scratches) in an image or video by using the information from the surrounding regions of the damage. Such techniques exploit knowledge of optical flows surrounding the damaged region(s) and hence depend on solutions of Partial Differential Equations such as the Navier-Stokes equation (Sochen & Fishelov, 2006), (Bertalmio, et al., 2001), (Fei, et al., 2008). On other hand, recently image inpainting based on sparse representation has been proposed for recovering the missing or damaged regions of an image (Jialun & Xianghong, 2013), (Shen, et al., 2009). The quality of the output image depends on the size and structure of the missing/damaged regions. If the region is large and lots of information is missing, then the inpainted image will be of low quality. The main difference between SR and inpainting is that the missing data in the case of SR is spread across all image rather than at certain regions. Hence SR technique is more suitable than inpainting for the situation the missing data spread randomly on all low-resolution image where interpolation technique uses to recover the missing pixels from available data points in the low-resolution image and enlarge the size for SR procedure.

### 2.3.1 Super Resolution Model

Super resolution is an inverse problem for the solution, in that one assumes that the LR images are the outcome of applying a set/sequence of image operators on an imagined HR image. It is necessary to formulate an observation mathematical model, that determines how the observed LR image(s) relate to (or are obtained from) the HR image. Consider a HR image $x$ of size $L_1M_1 \times L_2M_2$ with the parameters $L_1$& $L_2$ representing the down-sampling factors in the observation model, the horizontal and vertical directions respectively. Each observed LR image of size $M_1 \times M_2$ is represented by $\boldsymbol{y_K}$ where $\boldsymbol{K = 1, 2, \dots, p}$ and $\boldsymbol{p}$ is the total number of LR images. Obviously, the observed LR images are blurred and down-sampled images of the HR image $\boldsymbol{x}$. Mathematically the SR problem can therefore be modelled as a solution of:

$$\boldsymbol{y_K = SBx} \quad \text{For} \quad \boldsymbol{1 \leq K \leq p}$$

or

$$\boldsymbol{y = Ax} \tag{2-12}$$

where $\boldsymbol{B}$ and $\boldsymbol{S}$ represents blur and down-sampling matrix respectively. The process for constructing a super resolved image from multiple LR images involves three steps (Chaudhuri, 2001) as illustrated in Figure 2.5. The first step is registration of the LR images; second, interpolation and the third step is restoration. These steps can be implemented sequentially or simultaneously according to the reconstruction methods adopted.

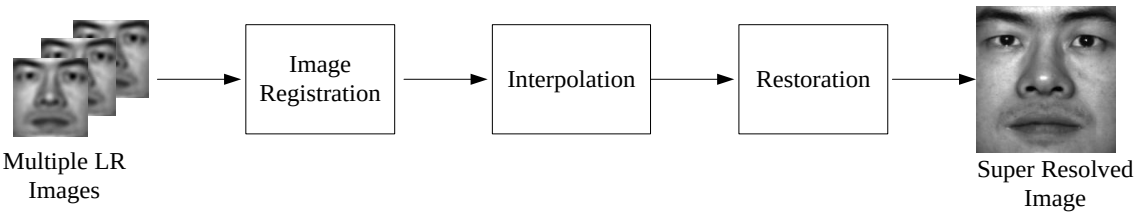



**Figure 2.5:** Three basic steps of super resolution image reconstruction.

The fundamental problem addressed in resolution enhancement, as presented in the above model, is an example of an inverse problem wherein the HR image is estimated from the degraded observations (LR images). Generally, such a reconstruction process is an ill-posed problem because of the insufficient number of LR images; ill conditioned registration where small variations in the observed images can cause large changes in the reconstruction; unknown blurring operator make the solution from the reconstruction constraint is not unique.

In mathematics different regularization approaches are proposed to stabilize the inversion of such an ill-posed problem, such as (Farsiu, et al., 2004), (Bannore, 2009) & (Zibetti, et al., 2008). Where the regularization term (i.e. $\|\boldsymbol{Qx}\|_2^2$ ) is often referred to as smoothing term, takes control of the ill-conditioned nature of the problem, and makes the solution more stable and as close as possible to the true solution. Therefore, the problem of SR in equation (2-12) is solved as:

$$\emptyset(\boldsymbol{x}) = \mathbf{min}[\|\boldsymbol{y} - \boldsymbol{Ax}\|_2^2 + \boldsymbol{\lambda}\|\boldsymbol{Qx}\|_2^2] \tag{2-13}$$

where $\boldsymbol{\lambda} > 0$ is the regularization parameter and $\boldsymbol{Q}$ is a regularization matrix where a high pass filter is a common choice for the operator $\boldsymbol{Q}$ and 2D Laplacian operator with ($\boldsymbol{Q}$= 4-neibourhood) has been used for the stabilization matrix (Fanaswala, 2009). The regularization parameter $\lambda$ controls the measure of smoothness in the final solution of equation (2-13) and it is choosing best suited to the particular application in which it is involved. Where a large value of $\boldsymbol{\lambda}$ affects the reconstruction output image and leads to a smoother image and some information may be lost. While too small values of $\boldsymbol{\lambda}$ lead to an unstable solution and make the approximate solution far from converging with the true solution. Hence, there needs to be a proper balance of smoothness and preservation of information when regularization is implemented. The regularization parameter $\boldsymbol{\lambda}$ could be either chosen manually (i.e. by visual inspection) or automatically, such as $\boldsymbol{\lambda}$-curve. In $\boldsymbol{\lambda}$-curve method the error between the original test and the reconstruction HR image is requiring (see (Fanaswala, 2009) and (Bannore, 2009)). However, the performance of regularization approaches degrades rapidly when the desired magnification factor is large or the number of the input LR images is small, which make the results overly smooth and lacking important high-frequency details.

### 2.3.2 Existing SR Techniques

Super resolution has been extensively studied and a lot of research is available in the literature, where over the last few years numerous SR methods have been developed and applied to a variety of image classes with differing degrees of success. Reconstruction-based iterative algorithm has proposed to reconstruct a HR image from multiple LR images (e.g. (Irani & Peleg, 1991), (Irani & Peleg, 1993) & (Peleg, et al., 1987)).

Irani and Peleg (Irani & Peleg, 1991) proposed the iterative back-projection (IBP) technique that seems to extend and generalise methods originally developed for computer-aided tomog-

raphy. Initially, more than one observed LR image are input to the proposed SR method, from which an initial up-sampled HR image $X^0$ is created by interleaving and interpolation. At the n-th iteration the SR procedure generates a set of simulated LR images (e.g. by down sampling) from the current HR image $\boldsymbol{X^n}$and then the HR image is updated until the error between the simulated LR images and the observed LR images is minimized. The updates are carried out by passing the error through a back-projection operator and this method can be expressed by the following equation:-

$$\boldsymbol{X^{n+1} = X^n + A^{bp} \sum_{i=1}^{K} (LR_i - LR_i^{sim})} \quad (2\text{-}14)$$

where $\boldsymbol{A^{bp}}$ is the back-project operator, $\boldsymbol{K}$ is the number of LR images and $\boldsymbol{n}$ is the integer's number. In their later work (Irani & Peleg, 1993), the iterative method applied for enhancing image frames using motion information when part of a tracked object was occluded in some frames and appears in others. The authors found that fusing information on tracked objects from number of registered frames enables reconstruction of occluded regions and improves image resolution. The iterative method will be described in greater detail in Subsection 2.3.3, where it is used to generate super resolved images from only two input LR images and the interpolation technique used to fill the missing pixels in the initial image $X^n$.

Another exiting category of SR approaches is non-iterative algorithms that generally have lower complexity and have expanded on image interpolation. For example, Chu et al. (Chu, et al., 2009) presented gradient-based adaptive interpolation method that depends on the local gradient of the original image and on the Euclidean distances between the interpolated pixel and its neighboring pixels (with size $5 \times 5$). The distances have been used to find three nearest pixels around the interpolated one after rank it from the closest to the farthest. Local gradient has been evaluated by averaging two sobal masks of size 3 (one horizontal and one vertical). The authors utilized the frequancy domain registration algorithm (Vandewalle, et al., 2006) to estimate the motions of the LR images and map these images onto the HR grid. The gradient-based adaptive interpolation method is used to form an HR image from the grid image. The wiener restoration technique has been applied to reduce the effect of blurring and noise caused by the system. In (Luong, et al., 2006) proposed a non-local interpolation method for image interpolation. Here, the estimation of the unknown pixel value is not based on its local surrounding neighbourhood, but overall of the particular image, based on repetitive characters of the image (i.e. parts of the image repeat themselves on an ever-diminishing scale). This method is perfectly suitable to some applications such as text images and satellite

images. On the other hand, a discriminative super resolution (DSR) technique has been proposed (Zou & Yuen, 2010) to overcome the problem of very LR face images where the face region is smaller than $16 \times 16$ pixels by finding the relationship between the LR images and the corresponding HR training images. Then super resolve image recovered by applying the relationship operator on the input LR image.

### 2.3.3 Back-Projection Iterative Interpolation Super Resolution

Super Resolution techniques are used to overcome the problem of image resolution, where the back project iterative interpolation super resolution (IISR) method in the spatial domain is one of the most approaches used to enhance/reconstruct image resolution from the information in the LR images. The iterative method starts with an initial guess $\mathrm{X}^{(0)}$ of the output HR image and then the down-sampling process is simulated to generate new LR images $\left\{LR_k^{(0)}\right\}$ based on the initial guess of the model, which corresponds to the observed input images $LR_k$. These *LR* images are then compared with the observed ones and the error between them is used to improve the initial guess by projecting back each value. This algorithm can be used to reconstruct an HR image from $\boldsymbol{LR_1}$& $\boldsymbol{LR_2}$ images only, where each image contains different information, as well as displaying an inability to fill all the pixels in the HR image. These images are created from the original image *X* after blurring as:

$$\boldsymbol{LR_1}(\mathbf{x_1}, \mathbf{y_1}) = \mathbf{X}(\mathbf{2x}, \mathbf{2y}) \;\&\; \boldsymbol{LR_2}(\mathbf{x_2}, \mathbf{y_2}) = \mathbf{X}(\mathbf{2x-1}, \mathbf{2y-1}) \qquad (2\text{-}15)$$

The initial approximation for the HR output image is estimated by combining/fusing the two LR images over a finer grid $\widehat{H}$ and bi-cubic interpolation method used to interpolate the missing pixels in the finer grid to obtain a single blurred image $\widehat{H} = \mathrm{X}^{(0)}$ of higher spatial sampling rate. Fusing the image information from $\boldsymbol{LR_1}$and $\boldsymbol{LR_2}$ to generate the image $\widehat{H}$ is formulated as:

(1) Transfer each pixel$(\boldsymbol{x}, \boldsymbol{y})$ in the $\mathbf{LR_1}$ image $\boldsymbol{Tr}\big(\boldsymbol{LR_1}(\boldsymbol{x}, \boldsymbol{y})\big) = \widehat{\boldsymbol{H}}(\boldsymbol{x_H}, \boldsymbol{y_H})$ by:-

$$\boldsymbol{x_H} = \mathbf{2}\boldsymbol{x}, \qquad \boldsymbol{y_H} = \mathbf{2}\boldsymbol{y}$$

(2) Transfer each pixel $(\boldsymbol{t}, \boldsymbol{w})$ in the $\mathbf{LR_2}$ image $\boldsymbol{Tr}\big(\boldsymbol{LR_2}(\boldsymbol{t}, \boldsymbol{w})\big) = \widehat{\boldsymbol{H}}(\boldsymbol{x_H}, \boldsymbol{y_H})$ by:-

$$\boldsymbol{x_H} = \mathbf{2}\boldsymbol{t} - \mathbf{1}, \qquad \boldsymbol{y_H} = \mathbf{2}\boldsymbol{w} - \mathbf{1}$$

The iterative step is then applied to remove possible artefacts from the initial approximation. The imaging process is simulated to obtain LR images by taking even and odd valued indices

from the initial approximation after blurring the initial image. The difference images between the input LR images and the simulated LR images are combined, leaving the other pixels equal to zero to obtain an error image of the same size as the original image or initial approximation. Then the wiener restoration technique is used to reduce the effect of blurring and noise. Finally, the de-blurred error image is used to improve the initial image, which yields an HR image $X^{(1)}$ and the new image is input to the next iteration cycle. The iterative update scheme of the output HR image can be expressed by

$$\boldsymbol{X^{(n+1)} = X^{(n)} + \left(\left(LR_k - LR_k^{(n)}\right) \uparrow s\right) * P} \qquad \boldsymbol{n = 0, 1, 2, \ldots, t.}$$

where $\boldsymbol{k}$ is the number of the $\boldsymbol{LR}$ images,↑ denotes up sampling by a factor $\boldsymbol{s}$ and $\boldsymbol{P}$ is a back-projection restoration filter. The SR scheme terminates either when the energy of the error term $\left(\boldsymbol{LR_k - LR_k^{(n)}}\right)$ is reduced below a certain threshold or the number of iterations reaches a fixed maximum number, or by minimizing the mean square error between hypothetical and reconstructed images.

Generally, the iterative interpolation method is used to improve the information in the low-resolution images. However, the question arising in this method can be used in the frequency domain for improving image resolution. In Chapter 4, we will test the viability of the iterative method to modify the wavelet subband of the input images.

### 2.3.4 Compressive Sensing Approach to SR

Signal and image reconstruction has been shown to benefit from the recently introduced field of ***compressive sensing*** and underpins the suitability of an efficient innovative category of SR approaches. Compressive sensing is a newly emerging mathematical theory that relaxes the strict conditions imposed by the S-N sampling theory for the recovery of sparse signals. It provides a different perspective in solving a set of linear equations when the number of rows is much less than the number of columns (i.e. underdetermined linear systems of equations). Underdetermined linear systems are either inconsistent or have infinitely many solutions that belong to an affine space. In the latter case, if we are seeking optimised solutions (such as sparse solutions or solutions that have the least square modulas) then we know that a unique solution exists and can be found. The least square method is a well established algorithm to find a unique solution vector that has the minimal Eucledian length but is ill-conditioned.

Moreover, our SR problem can be formulated in terms of sparse solutions to equations like (2-12). The corresponding $l_1$-minimization problem is known to be efficient, and in some cases this solution is also sparse. Compressive sensing is about the conditions on underdetermined measurement matricies (referred to in the literature as dictionaries), under which one can find the unique sparse solution of the underdetermint system, and efficient algorithms by which we can reconstruct the solution.

The compressive sensing topic exploits the empirical observation that many types of signals and images can be quite accurately approximated by sparse expansion in terms of suitable bases, by a relatively small number of non-zero coefficients. The recent advances in compressive sensing theory and the development of efficient $l_1$-minimization procedures to obtain sparse solutions for certain underdetermined linear systems, has led to the emergence of new SR schemes to recover high quality super resolved images from a single LR degraded image. In the past few years, many people have been using the new theory of sampling on single image SR problems to reconstruct an HR image. Yang et al in (Yang, et al., 2008) & (Yang, et al., 2010) proposed a method to reconstruct super resolved images using a pair of overcomplete dictionaries in which columns are constructed, through a learning process, from a number of randomly selected patches of high and low resolution tarining datasets. In Chapter 4 we will describe this method in more details, and subsequently will invetigate and test its performance in our non-training based system for face recognition in uncontrolled conditions. In (Zeyde, et al., 2010) and (Studer, 2010) the authors presented the K-means singular value decomposition (K-SVD) for generating learning dictionaries to scale-up LR images using sparse coding. This is an iterative algorithm that alternates between a sparse coding of the training samples for a current dictionary and dictionary update step, to produce dictionaries that better fit the data. The K-SVD algorithm performs a singular value decomposition (SVD) for each of the K different sub-matrices; hence the name of K-SVD (K is used as the number of columns in the dictionary; and SVD will be defined in the following chapter).

Bilgazyev et al. (Bilgazyev, et al., 2011) proposed an SR algorithm based on sparse representations, where sparse representation of the input LR image is computed using the dictionary built for LR images, and the HR image of high frequency components is estimated using the given sparse representation with respect to the HR dictionary. The estimated high frequency components of the HR image are then added to the LR input image to create an SR image. In this approach, the high and low dictionaries are prepared using dual tree complex

wavelet transforms from a set of training images. In (Xiaoqing, 2011) a two stage SR technique has been developed to reconstruct an HR image from a sequence of LR images. The first stage uses sparse representation to magnify each input LR image based on the learning dictionary. In the second stage, the output HR image is obtained by fusing the intermediate HR image sequence based on projection onto the convex sets (POCS) method.

An important concern and focus of CS research is the recovery of a sparse signal from a relatively small number of measurements of the signal. Understanding conditions of the underdetermind matrices which are identified in the CS literature will help to develop/or design novel dectionaries for recovering an SR image where these dictionaries satisfy statistical properties of the new theory of sampling. In the following chapter we will focus on the mathematical conditions of CS dictionaries and introduce some examples of different random dictionaries which are satisfy the well known property needed for sparse signal recovery.

# Chapter 3

## THE MATHEMATICAL CONCEPT OF COMPRESSIVE SENSING

Over the past few years has arisen significant interest in sparse representation of signals, in which this representation captures useful characteristics of the signal from the sensing matrix, which is known as the ***dictionary***. The central challenge for sparse recovery is the construction of preferably non-adaptive, relatively small numbers of linear measurements that can guarantee the reconstruction of a sparse or approximately sparse signal. Such a set of linear measurements are represented by rows of an overcomplete dictionary. This chapter is primarily concerned with studying the mathematical properties of the underdetermined dictionary that are relevant to the recovery of a sparse image from a down sampled degraded version of images.

This chapter first presents (Section 3.1) a very brief, but purposefully sufficient, historical introduction to the mathematical concepts of compressive sensing (CS), as a natural advanced extension of Harmonic and non-Harmonic Fourier and wavelet analysis of image signals that adopts an innovative approach to sensing and sampling which fits well with the aims of compression. Moreover, we discuss the obvious links to classical problems in linear algebra relating to finding the sparsest solution of underdetermined systems of equations formed by the action of overcomplete dictionaries. In Section 3.2, we review the necessary and/or sufficient structural conditions that CS dictionaries are expected to satisfy in relation to the task of recovering a sparse vector/signal using $L_1$-minimization. Furthermore, a variety of random measurement matrices known to satisfy the Restricted Isometry Property will be briefly described in section 3.3. Finally, in section 3.4 the summary and conclusions will be covered.

### 3.1 A Brief History of Compressive Sensing

Compressive Sensing, also known as sparse recovery, is a novel paradigm of signal sampling that emerged recently; but its roots can be traced back more than two centuries through the

mathematical field of Harmonic and non-Harmonic Analysis. Basically, CS relies on the main common assumption, adopted for image/signal compression, that bounded signals and images can be well approximated by sparse expansion in terms of a suitable set of atoms (i.e. signal building blocks). The new CS theory of sampling greatly relaxes the stringent requirements of the classical Shannon-Nyquist Theorem, which stipulates that perfect reconstruction of a signal require sampling at a rate below the Nyquist rate. Moreover, the two frameworks differ in the manner in which they deal with *signal recovery*. In the classical sampling framework signal recovery is achieved through a linear interpolation process, whilst in CS the signal recovery is typically achieved using the optimization method. Here we shall give a very brief historical account of how CS emerged.

Over the centuries, mathematical tools and transforms have been developed and refined to enable the removal of redundancies in signals, provide concise representation of these signals, and facilitate efficient processing and analysis of certain types of signals. In fact, the mathematical fields of functional and numerical analysis is dedicated to, among other things, investigating the properties, capabilities and limitations of such transforms in terms of signal representation/expansion. Examples of such tools include Fourier transforms, discrete Cosine transforms and wavelet transforms. The Fourier basis $\{\phi_n\}_{n=0}^{N-1}$ describes a signal $x \epsilon \mathbb{R}^n$ in terms of its global frequency content; by projecting it onto *K* lowest frequency, atoms (base vectors). However, the lack of localization makes it difficult to represent discontinuities because Fourier atoms do not have bounded support (i.e. window function with good localization in time and frequency domain). In 1946 (Gabor, 1946) Gabor resolved this problem by introducing a new approach to analysing signals, in which time and frequency play symmetrical parts, where the information conveyed by a frequency band in a given time-interval can be analysed in various ways. This has led to the emergence of wavelet transforms with ample classes of wavelet atoms (called wavelet filter banks). Different classes of wavelet atoms can provide multi-resolution analysis of signals into their frequency subbands in a variety of ways, and allow perfect reconstruction of the signal in an efficient manner. Digital versions of the Fourier, Cosine and wavelet transforms are among the most common tools in signal and image processing and analysis.

In the last two centuries, classes of Fourier and wavelet atoms (also known as dictionaries) have been used for representing a signal, where the coefficients are computed as linear com-

binations of the given atoms. Each of these dictionaries is known to be suitable for specific purposes. Bases of Fourier atoms are well suited for analysing harmonic signals such as music, whereas bases of wavelet atoms are more suited for detecting discontinuities and small features such as points and noise. Therefore, these kinds of dictionaries are not well equipped for representing high-dimensional signal data that normally has complex structures. Many other dictionaries have been developed over the last few years, such as *ridgelets* and *curvelets* for other specific purposes, (Candes & Donoho, 2000), (Rubinstein, et al., 2010). It has been recognised that combining different bases of atoms into *overcomplete* dictionaries can provide richer multi-purpose signal analysis tools (Rubinstein, et al., 2010). Designing such dictionaries together with designing efficient algorithms to recover sparse signals from the observed effect of such dictionaries are the main concerns of research in compressive sensing.

There is an obvious and strong synergy between CS and Numerical Linear Algebra. In fact, the objective of designing CS-compliant dictionaries and developing recovery algorithms can be expressed in terms of the problem of solving underdetermined systems of equations, where the number of equations is smaller than the number of variables. In this case, the columns of the system's measurement matrix are the atoms of the corresponding CS-dictionaries. In linear algebra, we know that underdetermined systems of full row rank have infinitely many solutions forming a hyperplane, but if we are interested in optimizing such as the case that the solution is small (i.e. sparse) then a unique solution exists (Trefethen & Bau III, 1997), (Levy, 2010). This connection is at the core of applying CS theory to sparse signal recovery and signal/image analysis, and is exploited in this thesis for face recognition from degraded low-resolution images.

The link between compressive sensing and data/signal compression stems from the fact that. The latter exploits the fact that signals/images can be well approximated in terms of a suitable basis (e.g., DCT and wavelet atoms) by vectors that have a relatively large number of coordinates of small quantities and replacing these coordinates produces a sparse representation of the signal/image. For example, image/video compression tools (e.g. MPEG and JPEG) use DCT and wavelet transforms as the first step in obtaining approximately sparse representation of any input image or video frame. The main disadvantage of this traditional approach to compression is their asymmetric performance in the sense that decompressing the image/video is much faster than compressing incoming streams. This is primarily the result of the computationally expensive quantisation step, which aims to obtain a sparse version of the

transformed coefficients. This has raised the legitimate question of whether we can immediately obtain the sparse representation (i.e. the significant measurements) of the signal without sampling at high rates that are assumed necessary.

The term "Compressed Sensing" was first introduced by Donoho (Donoho, 2006), and emerged as a result of investigating the conditions under which this question can be answered positively and developing algorithms to recover the required sparse signal from a significantly smaller number of measurements of the signal than that stipulated by the Shannon-Nyquist Theorem. However, the objectives of CS theory can be traced back to the late 1940s and early 1950s when there developed a great interest in non-harmonic Fourier expansion of signals as well as the development of wavelet transforms that evolved from the work and observations of Gabor. These efforts have led to the emergence of the ***theory of finite frames***, introduced in 1952 by Duffin and Schaeffer (Rubinstein, et al., 2010), which is the mathematical theory in which CS is deep rooted. A Frame is an overcomplete matrix that has more atoms than the dimensions of the signal, suitable for studying some problems in the non-harmonic Fourier series. In 1986 and 1990 Daubechies (Daubechies, et al., 1986), (Daubechies, 1990) introduced the fundamental concept of frames and studied the wavelet transform from the angle of frame theory. Nowadays, frame theory provides an extensive framework for the analysis and decomposition of signals in a stable and redundant way, where finite frames are regarded as the best generalization of orthonormal bases. As mentioned above, representations of signals with orthogonal dictionaries were prevalent for years in signal processing techniques, e.g. wavelet for compression. However, redundancy and over-completeness of frames provide more flexibility as well as robustness than the orthonormal bases. Here, the frame constructed by the redundant/overcomplete system is used to fit a particular problem in a manner not possible by a set of linearly independent vectors. Many interesting and useful dictionaries form *tight frames*, which ensure the representation of signals as a linear combination of the atoms (i.e. the inner product of the signal with the dictionary columns).

Over the last decade compressive sensing has become a key concept in various areas of applied mathematics and computer science for solving a class of underdetermined problems in computer vision. The list of applications of CS is growing fast and includes image/signal processing for improve image resolution, medical imaging and many others. In this thesis, we shall investigate the use of CS in dealing with the computer vision challenge of super resolu-

tion applied for the purpose dealing with the problem of face recognition in uncontrolled conditions where the captured face image is of low-resolution and of degraded quality.

### 3.1.1 Orthonormal Basis and Frames

This section provides a brief review of some of the key concepts in mathematics that were fundamental in developing the CS theory. Throughout this section, $\mathbb{R}^m$ stands for the real vector space of m-dimensional arrays of real numbers with coordinate wise addition of vectors and coordinate wise multiplication by scalars.

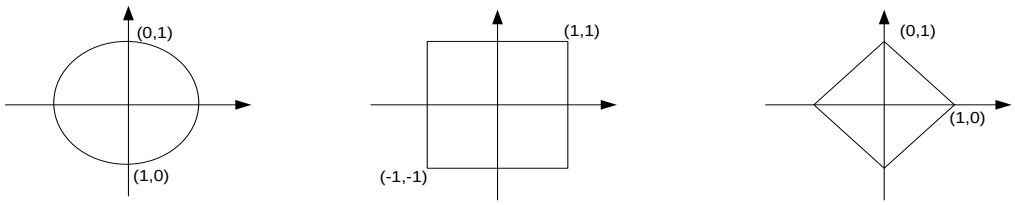
The inner product of two vectors $\boldsymbol{x} = (\boldsymbol{x_1}, \dots, \boldsymbol{x_m})$ and $\boldsymbol{y} = (\boldsymbol{y_1}, \dots, \boldsymbol{y_m})$ is defined as:

$$< \boldsymbol{x}, \boldsymbol{y} > = \sum_{\boldsymbol{i=1}}^{\boldsymbol{m}} \boldsymbol{x_i y_i}$$

For $p\epsilon[1, \infty)$, the $l_p$-norm of a vector $\boldsymbol{x}\boldsymbol{\epsilon}\mathbb{R}^{\boldsymbol{m}}$ is defined as:

$$\|\boldsymbol{x}\|_{\boldsymbol{p}} = \begin{cases} \left(\sum_{\boldsymbol{i=1}}^{\boldsymbol{m}} |\boldsymbol{x_i}|^{\boldsymbol{p}}\right)^{\boldsymbol{1/p}} & \boldsymbol{p\epsilon}[\boldsymbol{1}, \infty) \\ \max\limits_{\boldsymbol{i=1,2,..m}} |\boldsymbol{x_i}| & \boldsymbol{p} = \infty \end{cases}$$

The $l_2$-norm is known as the Euclidean norm, whereas the $l_1$-norm is referred to as the Manhattan norm or the city-block norm. Note that norms can be used to define a distance function on the points in $\mathbb{R}^m$. However, not every distance function is associated with a norm. Different norms defined on an m-dimensional space $\mathbb{R}^m$, endows $\mathbb{R}^m$ with a different geometry structure complementing its algebraic structure. For example, geometrically the $l_p$-norms on the space $\mathbb{R}^2$ associate different spatial shapes with unit the $l_p$balls $\{x\epsilon\, \mathbb{R}^2, \|x\|_p \leq 1\}$, see Figure 3.1, below:




$\|x\|_2 \leq 1$ $\|x\|_\infty \leq 1$ $\|x\|_1 \leq 1$

**Figure 3.1:** Unit cells in $\mathbb{R}^2$ for three norms

Compressive sensing relies on two fundamental principles; sparsity (sparse signal) and on an overcomplete dictionary. The basic notion of ***sparsity*** states that a vector has at most $k$ non-zero coefficients and it is measured by the $l_0$-norm. Thus the $l_0$-norm can be defined as:

**Definition** 3.1 **($l_0$-norm):** Let $\boldsymbol{x} = (\boldsymbol{x_i})_{\boldsymbol{i=1}}^{\boldsymbol{m}} \boldsymbol{\epsilon} \mathbb{R}^{\boldsymbol{m}}$ be a vector, then $l_0$-norm for $\boldsymbol{x}$ is

$$\|\boldsymbol{x}\|_{\boldsymbol{0}} = \sum_{\boldsymbol{i=1}}^{\boldsymbol{m}} [\boldsymbol{x_i}]$$

where $[\cdot]: \mathbb{R} \to \mathbb{R}$ is a function and

$$[\boldsymbol{x}] = \begin{cases} \boldsymbol{1} & \boldsymbol{if\ x \neq 0} \\ \boldsymbol{0} & \boldsymbol{if\ x = 0} \end{cases}$$

This norm does not really constitute a mathematical norm due to the fact that it violates one of the essential axioms in the definition of norms. In fact, $\|\alpha v\| = |\alpha|\|v\|$ when $\alpha$ is a negative scalar and v is an element of a vector space $V$ does not hold. Nevertheless, we shall abuse the language and refer to it as the $l_0$-norm due to the fact that it satisfies the property of sub-linearity as in the following lemma which will be used later to prove theorem (3.3).

**Lemma** 3.1**(Sub-linearity of $l_0$-norm):** For any vectors $\boldsymbol{x}$ and $\boldsymbol{y}$ of length-$m$ in the space $\mathbb{R}$

$$\|\boldsymbol{x} + \boldsymbol{y}\|_{\boldsymbol{0}} \leq \|\boldsymbol{x}\|_{\boldsymbol{0}} + \|\boldsymbol{y}\|_{\boldsymbol{0}}$$

**Definition** 3.2 **($k$-sparse):** A vector $\boldsymbol{x} = \begin{bmatrix} \boldsymbol{x_1} \\ \boldsymbol{x_2} \\ \vdots \\ \boldsymbol{x_m} \end{bmatrix} \boldsymbol{\epsilon} \mathbb{R}^{\boldsymbol{m}}$ is called $k$-sparse, if

$$\|\boldsymbol{x}\|_{\boldsymbol{0}} = |\{\boldsymbol{i}: \boldsymbol{x_i} \neq \boldsymbol{0}\}| \leq \boldsymbol{k}$$

In addition, let

$$\boldsymbol{\Sigma_k} = \{\boldsymbol{x}: \|\boldsymbol{x}\|_{\boldsymbol{0}} \leq \boldsymbol{k}\}$$

A set of vectors $\boldsymbol{\Phi} = \{\boldsymbol{\varphi_1}, \boldsymbol{\varphi_2}, \dots, \boldsymbol{\varphi_m}\}$ is called a ***basis*** for $\mathbb{R}^m$ if the vectors in the set span $\mathbb{R}^m$ and are linearly independent. This implies that each vector $\boldsymbol{x} = \begin{bmatrix} \boldsymbol{x_1} \\ \vdots \\ \boldsymbol{x_m} \end{bmatrix}$ in the space has a unique representation as a linear combination of these basis vectors as

$$\boldsymbol{x} = \boldsymbol{c_1}\ \boldsymbol{\varphi_1} + \boldsymbol{c_2}\boldsymbol{\varphi_2} + \cdots + \boldsymbol{c_m}\boldsymbol{\varphi_m}$$

or

$$\boldsymbol{x} = \boldsymbol{\Phi c} \quad \text{Where} \quad \boldsymbol{c} = \{\boldsymbol{c_i}\}_{\boldsymbol{i=1}}^{\boldsymbol{m}} \tag{3-1}$$

An important special case of a basis is an orthonormal basis, defined as a set of vectors $\{\varphi_i\}_{i=1}^{m}$ satisfying

$$< \boldsymbol{\varphi_i}, \boldsymbol{\varphi_j} > = \begin{cases} \mathbf{1} & \boldsymbol{i = j} \\ \mathbf{0} & \boldsymbol{i \neq j} \end{cases}$$

and $$\|\boldsymbol{\varphi_i}\|_2 = \mathbf{1} \qquad \text{for} \quad \boldsymbol{i} = \mathbf{1}, \dots, \boldsymbol{m}$$

A ***frame*** generalises the concept of basis, in that it is spanning set of vectors with an extra property that is relevant to CS conditions on signal recovery. Note that a spanning set of vectors that has more vectors than the dimension of the spanned vector space generally involves redundancy. While each vector in a vector space has a unique representation as a linear combination of vectors in a basis, it will have infinitely many representations with respect to a frame of vectors. In general, the frame can be defined as:

**Definition** 3.3 **(Frame):** A family of vectors $(\boldsymbol{\varphi_i})_{\boldsymbol{i=1}}^{\boldsymbol{n}}$ in $\mathbb{R}^m$ or $\mathbb{C}^m$, corresponding to a matrix $\boldsymbol{\Phi} \epsilon \mathbb{R}^{m \times n}$ ,where $m < n$ is called a frame, if there exist constants $\mathbf{0} < \boldsymbol{A_{Lb}} \leq \boldsymbol{B_{Ub}} < \infty$ such that for all $\boldsymbol{x} \boldsymbol{\epsilon} \mathbb{R}^{\boldsymbol{m}}$;

$$\boldsymbol{A_{Lb}}\|\boldsymbol{x}\|_2 \leq \|\sum_{\boldsymbol{i=1}}^{\boldsymbol{n}} \langle \boldsymbol{x}, \boldsymbol{\varphi_i} \rangle\|_2 \leq \boldsymbol{B_{Ub}}\|\boldsymbol{x}\|_2 \tag{3-2}$$

Equivalently,

$$\boldsymbol{A_{Lb}}\|\boldsymbol{x}\|_2 \leq \|\boldsymbol{\Phi^T x}\|_2 \leq \boldsymbol{B_{Ub}}\|\boldsymbol{x}\|_2$$

The constants $\boldsymbol{A_{Lb}}$ and $\boldsymbol{B_{Ub}}$ are the lower and upper boundaries for the frame, respectively. The condition $\boldsymbol{A_{Lb}} > 0$ implies that the rows of $\boldsymbol{\Phi}$ must be linearly independent. If the frame boundaries are chosen as $\boldsymbol{A_{Lb}} = \boldsymbol{B_{Ub}}$ then the frame is called Tight frame (TF) and

$$\boldsymbol{A_{Lb}}\|\boldsymbol{x}\|_2 = \left\| \sum_{\boldsymbol{i=1}}^{\boldsymbol{n}} \langle \boldsymbol{x}, \boldsymbol{\varphi_i} \rangle \right\|_2$$

A tight frame $\boldsymbol{\Phi} = (\boldsymbol{\varphi_i})_{\boldsymbol{i=1}}^{\boldsymbol{n}}$for which $\boldsymbol{A_{Lb}} = \boldsymbol{B_{Ub}} = \mathbf{1}$ is called a Parseval frame (PF). In other cases, the underdetermine matrix $\boldsymbol{\Phi}$ is called Equal-norm frame (ENF) if all the elements

have the same norm i.e. $\|\boldsymbol{\varphi_i}\| = \boldsymbol{c}$, for all $\boldsymbol{i} = \boldsymbol{1}, \boldsymbol{2}, .., \boldsymbol{n}$ & $\boldsymbol{c} > 0$. However, if $\boldsymbol{c} = \boldsymbol{1}$, then the frame is called unit-norm (UN).

The research area of finite frames $(\varphi_i)_{i=1}^n$ has a vast number of applications. The coefficients vector of $\sum_{i=1}^n \langle x, \varphi_i \rangle$ can be utilized for transmission$x$ by the operator $x \longmapsto \sum_{i=1}^n \langle x, \varphi_i \rangle$ i.e. it maps the signal into a higher dimensional space, or utilizes for edge detection of an image $x$, or for recovery of missing data. Moreover, the finite frames have been developed and used as codes for erasures and additive noise. The noise and erasures are one of the most common problems of signal transmissions and these types of errors typically occur when analogue signals are transmitted in an unreliable environment. The general model for frame-coded transmissions contains three parts; linear encoding of a vector in terms of its frame coefficients, transmission, and reconstruction. Frames have been used as block codes, which replace the coefficients of a vector with respect to an orthonormal basis by a larger number of linear coefficients in the expansion with a stable spanning set.

On the other hand, finite frames play a central role in the design and analysis of sparse representation, where by using suitable chosen representation systems $\Phi = (\varphi_i)_{i=1}^n$allows us to find the sparse sequences $(c_i)_{i=1}^n$with small numbers of non-zero coefficients and to expand the data $x$ by $x = \sum_{i=1}^n c_i \, \varphi_i$. Therefore, if the frame $\Phi$ satisfies the unique representation property (Casazza, 2013) of order $2k$ then $k$-sparse signal $x$ can be uniquely recovered. The unique representation property can be defined as the following:

**Definition** (3.4) (**Unique Representation Property**): A frame $\Phi = (\varphi_i)_{i=1}^n$ is said to have the unique representation property of order $k$ if any $k$ frame elements of $\Phi$ are linearly independent.

Finally, we consider simple examples of frames. Let $\Phi \epsilon \mathbb{R}^{m \times m}$ be a standard basis matrix, and then the frame can be generated by adding zero-vectors of length $m$ to the basis system $\Phi$ to obtain a finite frame $\widehat{\Phi}$ of size $m \times n$, where $m < n$ with equal boundaries,

$$\widehat{\Phi} = \begin{bmatrix} 1 & 0 & 0 & 0 & \cdots & 0 & 0 \\ 0 & 0 & 1 & 0 & \cdots & 0 & 0 \\ 0 & 0 & 0 & 0 & \cdots & 0 & 0 \\ & & \vdots & & \ddots & \vdots & \\ 0 & 0 & 0 & 0 & \ldots & 1 & 0 \end{bmatrix}$$

This is a Parseval frame and $\left\|\sum_{i=1}^{n}\langle x,\widehat{\Phi}_i\rangle\right\|_2 = \|x\|_2$ for all vectors in the space $\mathbb{R}^m$.

Other examples can be constructed by adding multiple copies of vectors in a given basis, e.g.

$$\Phi_1 = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & \cdots & 0 & \cdots & 0 \\ 0 & \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} & 0 & 0 & 0 & \cdots & 0 & \cdots & 0 \\ 0 & 0 & 0 & \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{3}} & \cdots & 0 & \cdots & 0 \\ & & & & \vdots & & \ddots & \vdots & & \\ 0 & 0 & 0 & 0 & 0 & 0 & \cdots & \frac{1}{\sqrt{m}} & \cdots & \frac{1}{\sqrt{m}} \end{bmatrix} \quad \text{Or} \quad \Phi_2 = \begin{bmatrix} 1 & 1 & 0 & 0 & \cdots & 0 & 0 \\ 0 & 0 & 1 & 1 & \cdots & 0 & 0 \\ 0 & 0 & 0 & 0 & \cdots & 0 & 0 \\ & & \vdots & & \ddots & \vdots & \\ 0 & 0 & 0 & 0 & \cdots & 1 & 1 \end{bmatrix}$$

The matrix $\Phi_1$is a Parseval frame in which vector norms converge to zero as $m$ converges to infinity, while $\Phi_2$ is 2-tight frame where each vector $e_i$ appears twice.

Recall that the main premise of CS theory and its applications is concerned with signals that are either inherently sparse or can be made sparse using some reversible transformations, and the objective is to recover sparse signals from a small number of linear measurements (coding) of their samples. Ideally, the measurements should be non-adaptive and requiring no prior knowledge of all the signal samples (all pixel values in the case of images). The principle of sparse coding (sparse representation) assumes that a natural signal can be compactly expressed/and represented efficiently as a linear combination of the column vectors of a prespecified dictionary where most of the linear coefficients are zero. Therefore, the fundamental principle of CS relies on frame-like matrices, which are called ***Dictionaries*** and again can be defined (Rubinstein, et al., 2010) as a generalization of vector space basis. Therefore, CS dictionaries are represented by overcomplete $m \times n$ matrices, where $m \ll n$ and whose columns form a pool of $\mathbb{R}^m$ bases that span the entire vector space. This means that any vector in $\mathbb{R}^n$ can have multiple representations in terms of different bases in $\mathbb{R}^m$. Since $m \ll n$, then the multiplication of CS dictionaries by vectors in $\mathbb{R}^n$ yields an underdetermined system of linear equations where the number of rows is significantly less than the number of columns.
A system of $m$ linear equations with $n$ unknowns can be defined in general as:

$$\begin{aligned} \boldsymbol{y_1} &= \boldsymbol{d_{11}x_1} + \boldsymbol{d_{12}x_2} + \cdots + \boldsymbol{d_{1n}x_n} \\ \boldsymbol{y_2} &= \boldsymbol{d_{21}x_1} + \boldsymbol{d_{22}x_2} + \cdots + \boldsymbol{d_{2n}x_n} \\ &\vdots \\ \boldsymbol{y_m} &= \boldsymbol{d_{m1}x_1} + \boldsymbol{d_{m2}x_2} + \cdots + \boldsymbol{d_{mn}x_n} \end{aligned}$$

Alternatively, the system can also be written in the form of matrices as follows:

$$\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_m \end{bmatrix} = \begin{bmatrix} d_{11} & d_{12} & \dots & d_{1n} \\ d_{21} & d_{22} & \dots & d_{2n} \\ & & \vdots & \\ d_{m1} & d_{m2} & \dots & d_{mn} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}$$

or

$$\boldsymbol{y = Dx} \tag{3-3}$$

The assumption here, is that we only have the observed output vector $y$ and we are interested in recovering a unique solution which is the sparsest, i.e. has the smallest (minimum) $l_0$-norm.

In general, solving a system of linear equations of the form (3-3) is dependent on the relationship between $m$ and $n$. When $m = n$, and if $D$ is invertible then the equation (3-3) has the unique solution $x = D^{-1}y$. If $n < m$ then the system is redundant and may not have a solution. However, in the case of $m < n$ (i.e. our case of the system being underdetermined) then if a solution exists then there would be infinitely many solutions belonging to a hyperplane of dimension $n - m$.

**Example1.**

In this example we illustrate that the underdetermined system of few equations than variables has no unique solution. Let the set of two equations and four variables as the form:

$$\boldsymbol{x + y - 3z + w = 1}$$
$$\boldsymbol{x - 2y + 3z - w = 5}$$

From this system, we can get one equation by addition:

$$\boldsymbol{2x - y = 6} \tag{3-4}$$

Which is equivalent to $\boldsymbol{x = \frac{1}{2}\, y + 3}$

To solve this equation two assumptions exist. First, for example let $y = 0$ then the variable $x$ is equal to 3 and the two variables $z$ and $w$ can be determined by substituting $x\ and\ y$ in the first\or second equation to the above system to get the equation:

$$\boldsymbol{-3z + w = -2}$$

If $z = 0$ then $w = -2$. Hence, the first sparse solution is of the form $(3,0,0,-2)$. However, if the variable $w = 0$ then $z = 2/3$ and that leads to the second sparse solution which is equal to $(3,0,2/3,0)$.

The second assumption to the equation (3-4) is that if $x = 0$ then $y = -6$ as well as the two variables $z$ and $w$ are equals 0 and 7 respectively after substituting $x$ and $y$ in the first equation of the system and assuming $z = 0$. Therefore, the third sparse solution is $(0, -6, 0, 7)$. However, if we assume $w = 0$ , $x = 0$, and $y = -6$ then $z = -7/3$ and the four sparse solution is equal to $(0, -6, 0, -7/3)$. Therefore, the above underdetermined $2 \times 4$ system of equations has 4-solutions of 2-sparse.

In this example, the equation of the hyperplane was easily found and it was easy to search for sparse solution(s). Note that we find more than one sparse solution, which is not ideal for CS applications.

The CS problem of finding the sparsest solution for the given vector $y$ and measurement matrix $D = \{d_1, d_2, \dots, d_n\} \epsilon \mathbb{R}^{m \times n}$ such that $y = Dx$ exactly or approximately can be reformulated as finding a vector $\hat{x} \epsilon \mathbb{R}^n$ with a minimum possible number of nonzero entries. That is

$$\hat{\boldsymbol{x}} = \boldsymbol{arg}\,\boldsymbol{\min}_{\boldsymbol{x}} \|\boldsymbol{x}\|_{\boldsymbol{0}} \ \boldsymbol{subject\ to}\ \boldsymbol{y} = \boldsymbol{Dx} \tag{3-5}$$

**Example 2**.

Consider the $2 \times 4$ underdetermined system of two equations as the form:

$$\begin{pmatrix} \mathbf{1} & \mathbf{0} & \mathbf{4} & \mathbf{6} \\ \mathbf{0} & \mathbf{1} & \mathbf{-4} & \mathbf{3} \end{pmatrix} \begin{pmatrix} \boldsymbol{x} \\ \boldsymbol{y} \\ \boldsymbol{z} \\ \boldsymbol{w} \end{pmatrix} = \begin{pmatrix} \mathbf{1} \\ \mathbf{-1} \end{pmatrix}$$

To solve this system, summation of the two equations generates one equation as:

$$\boldsymbol{x} + \boldsymbol{y} + \mathbf{9}\boldsymbol{w} = \mathbf{0}$$

and assume that two of these variables are equal to zero to find another variable. Therefore, in any cases $x, y,$ and $w$ are equal to zero. To find the value of the fourth variation $z$ in the above system, we substitute all of these variables in the first or second equation and get $z = 1/4$, hence the solution is (0,0,1/4,0).

Now if we assume $x, y$ and $z$ are equal zeros in the above linear system then we will get two values for the variable $w$ which are $w = 1/6$ and $w = -1/3$ i.e. any other assumption than $x = y = w = 0$ we cannot get a solution. Therefore, the above-underdetermined linear equation has a unique solution of 3-sparse which is equal $(0, 0\ , 1/4, 0)$.

Solving (3-5) is an optimization problem with an exponential computational complexity. It is an NP-hard problem because no-deterministic algorithm is known to solve the problem in polynomial time. However, the differences between the above $2 \times 4$ matrices in examples 1 and 2 indicate that imposing some conditions may guarantee the existence of unique solutions of (3-5). For example, one can observe that in example1 the last two columns are linearly dependent, whereas every pair of columns of the matrix $D$ in example 2 is linearly independent. Studying the necessary and/or sufficient conditions that dictionaries need to satisfy (3-3) in order to have a unique solution is at the heart of CS research. Moreover, even if we construct such matrices then we still need to deal with the equally difficult challenge of designing efficient algorithms that can be used to recover the unique sparse solution for high dimensional systems. In short, the questions are: first, under what conditions does the CS problem have a unique sparse solution? Second, is there a simple test to verify the condition(s**)**?. These issues will be discussed in the next section, but before that, we briefly discuss a common approach that mathematicians follow to circumvent NP-hardness.

Mathematicians more often than not work by analogy and try to learn from other similarly stated problems (in this case linear optimisation) that we know how to solve. The problem of finding the best line or curve fit for a set of observed experimental data is by far the most popular computational optimisation problem that has significant similarity to that of finding the sparsest solution of the underdetermined system of equation (3-3). In fact, the class of optimisation problems referred to here, includes the well known the $l_2$-minimisation ($P_2$)-problem:

$$\boldsymbol{\hat{x} = arg\,\min_x \|x\|_2 \;\; subject\; to\; y = Dx} \tag{3-6}$$

It is well known that when $x$ is assumed to be a small vector (e.g. its length is minimal) then the Least Square Method (LSM) can be used to solve (3-6). The solution is based on the use of the Moore-Penrose pseudo-inverse matrix $D^+ = D^T(DD^T)^{-1}$, (Alves & Hussein, 2012), i.e.

$$\boldsymbol{\hat{x} = D^T(DD^T)^{-1}y}$$

where $D^T$ is transposing the matrix $D$. However, the LSM is ill-conditioned, doesn't yield a sparse vector, and it is not desired in many applications. Note that the $l_2$ –minimisation solution $\hat{x}$ is the contact point with the Euclidean ball, as illustrated in Figure 3.2 below, which does not coincide with the actual sparse signal $x$.

An alternative optimisation model that has common formulation to the $l_0$-minimisation is provided by the use of the $l_1$-norm instead, i.e. the following $l_1$-minimisation problem:

$$\hat{\boldsymbol{x}} = \boldsymbol{arg}\,\boldsymbol{\min}_{\boldsymbol{x}}\|\boldsymbol{x}\|_{\boldsymbol{1}}\ \ \boldsymbol{subject\ to}\ \boldsymbol{y} = \boldsymbol{Dx} \tag{3-7}$$

This is a convex optimisation problem, which is amenable to linear programming. More importantly, it has been shown that subject to certain conditions on the dictionary the solution of the $l_1$-minimisation coincides with the sparsest solution. Note that the solution $\hat{x}$ by $l_1$-norm is coincide with the actual sparse signal $x$ differs than the solution by $l_2$-norm (see Figure 3.2 below). This is because the geometry of the Euclidean ball does not lend itself well to detecting sparsity.

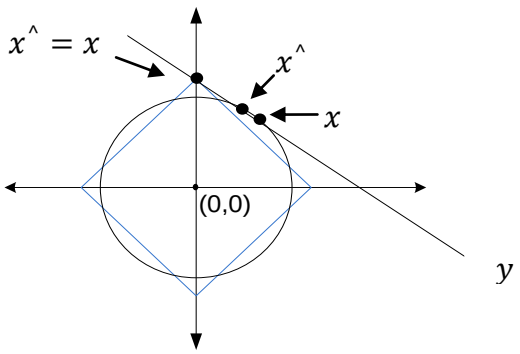



**Figure 3.2:** $\boldsymbol{l_1}$-minimization versus $\boldsymbol{l_2}$-minimization

The Linear Programming ***simplex*** method (Chen & Donoho, 1994) is a guaranteed algorithm to solve the problem (3-7) of sparse signal recovery from CS measurements. The simplex method starts from an initial basis $B$ consisting of $n$ linearly independent columns of $D$ for which the corresponding solution $B^{-1}y$ is feasible (non-negative). Then improves the current basis iteratively by, at each step, swapping one term in the basis for one term not in the basis, with the goal of improving the objective function. Finally, the solution is achieved when no improvement is possible in the iterative improvement of a basis. While $l_1$-minimization convex optimization technique is a powerful method for computing sparse representations, there is a variety of greedy/iterative methods for sparse signal recovery (e.g. see (Fornasier, 2010), (Eldar & Kutyniok, 2012), (Kutyniok, 2013), (Tropp & Wright, 2010)). The most commonly used iterative greedy approaches include Orthogonal Matching Pursuit (OMP) and Iterative Hard Thresholding (IHT).

## 3.2 Compressive Sensing Dictionaries and Algorithms

This section is devoted to the study of the main properties of CS-compliant dictionaries that enable the recovery of the unique solutions of the $l_1$- optimisation problem (3-7) which also solves (3-3). We assume that we are seeking $k$-sparse solutions $x$ with $k$ being sufficiently smaller than the number of columns of $D$, i.e. $x\epsilon\Sigma_k$. In general, the discussion presented earlier about CS dictionaries, or finite frames, is linked to how the columns of $D$ can provide different $k$-sparse representations of vectors in $\mathbb{R}^n$. Therefore, CS-compliance properties will be linked to linear combinations of $k$ columns of the dictionaries.

### 3.2.1 Sparkness of a matrix

The spark of a matrix $D\epsilon\mathbb{R}^{m\times n}$, was introduced by Donoho and Elad in (Donoho & Elad, 2003), and is defined as the minimum number of columns of $D$ that are linearly dependent. The matrix is said to have full spark if $spark\ (D) - 1 = rank\ (D)$ i.e. the matrix $D$ has a full full row rank. In (Kutyniok, 2013) the definition of spark is reformulated by using a null space $N(D) = ker(D) = \{x: Dx = 0\}$, i.e. as

$$\boldsymbol{spark(D) = min\{k: \mathcal{N}(D) \cap \Sigma_k \neq \{0\}\}} \tag{3-8}$$

Clearly, $1 < spark(D) \leq m + 1$.

Therefore, the value of spark is very informative and large values are evidently very useful because if the matrix $D$ has spark = $(m + 1)$, then there are no $m$-columns that are linearly dependent.

The following theorem provides the guarantee of recovery of the unique $k$-sparse solution of the underdetermined system.

**Theorem** 3.1 (Duarte & Eldar, 2011)**:** For any vector $y\epsilon\mathbb{R}^m$, there exists at most one sparse signal $x\epsilon\Sigma_k$ such that $y = Dx$ if and only if $\|x\|_0 < \text{spark}\ (D)/2$.

***Proof:*** Based on the proof of (Duarte & Eldar, 2011) the both directions will prove by contradiction.

(⇒) Suppose first that there exists a sparse signal $x$ such that $y = Dx$ for any $y\epsilon\mathbb{R}^m$ and assume $spark(D) \leq 2k$ where $k = \|x\|_0$. This means that there exists at most $2k$-columns that

are linearly independent. Thus from equation (3-8) there exists a vector $h\epsilon\mathcal{N}(D)$ such that $h$ is $2k$-sparse vector. Because $h\epsilon\Sigma_{2k}$ then we can write $h = x_1 - x_2$ where $x_1, x_2\epsilon\Sigma_k$ and $D(x_1 - x_2) = 0$ hence $Dx_1 = Dx_2$. Thus, there are two $k$-sparse solutions of $y = Dx$ and that contradicts with the expected result. Therefore, we must make spark $(D) > 2k$.

($\Leftarrow$) For the converse, let us assume that spark $(D) > 2k$ and suppose for some $y$ there exists $x_1, x_2\epsilon\Sigma_k$ such that $y = Dx_1 = Dx_2$ which implies $Dx_1 - Dx_2 = 0$ hence $D(x_1 - x_2) = 0$. Now let $h = x_1 - x_2$ then $Dh = 0$. Since spark $(D) > 2k$, then all sets of up to $2k$-columns of $D$ are linearly independent and $h = 0$ therefore $x_1 = x_2$. ■

This theorem provides an efficient strategy for constructing suitable CS dictionaries by combining/concatenating certain sets of $\mathbb{R}^m$ bases. However computing the spark of a dictionary is NP-hard. On the other hand, the absence of this property can be checked statistically by testing a randomly selected large set of $m$-columns for independence.

### 3.2.2 The Null Space Property (NSP)

Let $D$ be a $m \times n$ dictionary, and $k \leq m$ be a positive. The order $k$ NSP property of the dictionary $D$ is related to the size of any $k$ coordinates of the vectors in the null space of $D$, $\mathcal{N}(D) \equiv Ker(D) = \{x: Dx = 0\}$. We first need to explain some notations.

Let $\Omega \subseteq \{1,2,\dots,n\}$ be any subset of size $k$, we define the $k$-sparse version of a vector $x$ in $\mathbb{R}^n$, defined by $\Omega$ as the $n$-dimensional vector whose i-th coordinate is $x_i$ if $i \epsilon \Omega$ and 0 otherwise. .

For instance, let the vector $x = (x_1, x_2, \dots, x_n)$ and $\Omega = \{1,5,7,9\} \subseteq \{1,2,\dots,\mathrm{n}\}$ where $\mathrm{card}(\Omega) \leq 4$ then the vector $x_\Omega$ obtained from $x$ by:

$$\boldsymbol{x_\Omega = x_{\{1,5,7,9\}} = (x_1, 0, 0, 0, x_5, 0, x_7, 0, x_9, 0, 0, \dots)}$$

On the other hand, we obtain the vector $x_{\Omega^c}$ by replacing $x_i$ with 0 if $i \epsilon \Omega$ and keeping the others. For the above example:

$$\boldsymbol{x_{\Omega^c} = x_{\{2,3,4,6,8,10,11,\dots\}} = (0, x_2, x_3, x_4, 0, x_6, 0, x_8, 0, x_{10}, x_{11}, \dots)}$$

**Definition** (3-5). An $m \times n$ dictionary $D$ is said to satisfy the NSP of order $k$ if there exists a constant $C\epsilon(0,1)$ such that for each size $k$ set $\Omega \subseteq \{1,2,\dots,n\}$ and each vector $x\epsilon\mathcal{N}(D) \setminus \{0\}$,

$$\|\boldsymbol{x_{\Omega}}\|_1 \leq \boldsymbol{C}\|\boldsymbol{x_{\Omega^c}}\|_1 \tag{3-9}$$

An equivalent description of NSP is to say, if every $k$-columns of the overcomplete dictionary $D$ is linearly independent then every $(m/2)$-sparse $x$ can be recovered uniquely from $Dx$ by $l_1$-norm. This is a version of the so-called NSP, which is known to be necessary and sufficient for $l_1$-recovery.

**Lemma** 3.2: Let $D\epsilon\mathbb{R}^{m\times n}$ be an over-complete matrix and satisfies NSP property of order $k$ then every $k$-columns of $D$ are linearly independent.

***Proof (by contradiction).*** Assume there is a set of $k$-columns of $D$ that are linearly dependents.

Let, the $\boldsymbol{\Omega} = \{\boldsymbol{i_1}, \boldsymbol{i_2}, \dots, \boldsymbol{i_k}\} \subseteq \{\boldsymbol{1}, \boldsymbol{2}, \dots, \boldsymbol{n}\}$ is the index of the linearly dependent vectors $\boldsymbol{d_{i_1}}, \boldsymbol{d_{i_2}}, \dots, \boldsymbol{d_{i_k}}$ in the dictionary $D$ of size $m \times n$, where $(m < n)$.

By the definition of the linearly dependent vectors, there exists a scalars $\boldsymbol{x_{i_1}}, \boldsymbol{x_{i_2}}, \dots, \boldsymbol{x_{i_k}}$ not all zero such that the column vectors $\{\boldsymbol{d_{i_1}}, \boldsymbol{d_{i_2}}, \dots, \boldsymbol{d_{i_k}}\}$ in $D$ can be written as a linear combination as:

$$\boldsymbol{x_{i_1}d_{i_1}} + \boldsymbol{x_{i_2}d_{i_2}} + \cdots + \boldsymbol{x_{i_k}d_{i_k}} = \boldsymbol{0}$$

Now, let $x_{\Omega} = (\boldsymbol{x_{i_1}}, \boldsymbol{x_{i_2}}, \dots, \boldsymbol{x_{i_k}})$ and the vector $x_{\Omega}$ can be written as sparse vector of length n

$$\boldsymbol{x_{\Omega}} = (\boldsymbol{x_{i_1}}, \boldsymbol{x_{i_2}}, \dots, \boldsymbol{x_{i_k}}, \boldsymbol{0}, \boldsymbol{0}, \dots, \boldsymbol{0})$$

In addition

$$\boldsymbol{Dx_{\Omega}} = \boldsymbol{0}$$

However, $\boldsymbol{x_{\Omega}}$ do not satisfy the inequality (3-9). Hence, $D$ does not satisfy NSP property of order $k$.

■

**Remark.** Note that the matrix in Example 1, above, does not satisfy the condition in this theorem because the last two columns are linearly dependent. While, the matrix in Example 2 complies with this condition.

The following theorem provides a method for recovering the sparsest solution, which links the linear independence of certain numbers of columns to the uniqueness of sparse solutions of the $l_1$-minimization.

**Theorem** 3.2 (Fornasier & Rauhut, 2010): The $m \times n$ dictionary $D$ satisfies NSP of order $k$ if every $k$-sparse solution $x$ can be recovered by $l_1$-minimization.

**Theorem** 3.3**:** If every $2k$ columns of $D$ are linearly independent then $y = Dx$ has a unique $k$-sparse solution.

***Proof.*** Suppose that every $2k$ columns of $D$ are linearly independent and assume there are two $k$-sparse solutions $x_1$ and $x_2$ in $\mathbb{R}^m$ such that $y = Dx_1 = Dx_2$. Which implies that $D(x_1 - x_2) = 0$ and since $x_1$ & $x_2$ are both $k$-sparse solutions, then by lemma (3.1)

$$\|\boldsymbol{x_1} + (-\boldsymbol{x_2})\|_0 \leq \|\boldsymbol{x_1}\|_0 + (-\|\boldsymbol{x_2}\|_0) \leq \boldsymbol{k} + \boldsymbol{k} = \boldsymbol{2k}$$

Hence, $(x_1 - x_2)\epsilon\Sigma_{2k}$, so there is a linear dependence between $2k$ columns of $D$, and that contradicts with the assumption in the theorem.

■

Using this theorem to test for satisfying NSP, unfortunately, is an exhaustive search, which is not feasible. Therefore, NSP is rather restrictive and below we shall discuss a property that includes a larger space of dictionaries that meet the requirements of compressive sensing. However, a simple statistical test procedure can be used to show non-compliance with NSP. Simply, one can test a sufficiently large sample of $k$-column sub-matrices of $D$ for invariability. This statistical test can also be used as an indicator of compliance with the theorem if a sufficiently large sample shows positive response.

### 3.2.3 The Restricted Isometry Property

In 2006, Candes and Tao (Candes & Tao, 2006) introduced another CS related property which is less stringent than the NSP property. This is known as the Restricted Isometry Property (RIP) and it provides a sufficient condition for unique sparse solution using $l_1$-recovery. The $m \times n$ dictionary $D$, $m \ll n$ is said to satisfy the Restricted Isometry Property (RIP) of order $k$ if there is a constant $0 < \delta_k < 1$, such that for any $k$-sparse signal $x \in \mathbb{R}^n$:

$$(\mathbf{1}-\boldsymbol{\delta_k})\|\boldsymbol{x}\|_2^2 \leq \|\boldsymbol{Dx}\|_2^2 \leq (\mathbf{1}+\boldsymbol{\delta_k})\|\boldsymbol{x}\|_2^2 \tag{3-10}$$

The smallest such $\delta_k$ is called the Restricted Isometry Constant (RIC) of order $k$. Properties of sub-matrices by the columns of a dictionary that depend on the sparsity value of $k$ have been shown to indicate satisfy ability of RIP. Another way to expressing the RIP definition is by using condition number, in fact, it has been shown that if $D$ satisfies RIP of order $k$, then any $2k$-columns sub-matrix of $D$ must be well-conditioned (Baraniuk, et al., 2008) and (Rauhut, 2010). However, again checking this property for all $2k$-columns is computationally infeasible, as it requires exhaustive checks of all $\binom{n}{2k}$ sub-matrices. Therefore, a simpler way of ensuring recovery of sparse vector using $l_1$-minimization, and checking sufficient condition, is by computing the condition number of a sufficiently large set of randomly selected $2k$-submatrices from the CS matrix $D$.

The condition number of $D$ can be calculated using the singular value decomposition (SVD) of a matrix. Here, we present the definition of singular values of any matrix, which generalises the concept of eigenvalues to non-square matrices.

**Definition** 3.6 (**Singular Value Decomposition**):

Let $D\epsilon\mathbb{R}^{m\times n}$ a matrix, then the singular values $\sigma_1,\ \sigma_2, \dots, \sigma_n$ of $D$ are the square roots of the descending order sequence of the eigenvalues $\lambda_1, \lambda_2, \dots, \lambda_n$ of the Gramian square $n \times n$ matrix $M = D^T D$, i.e. $\sigma_i = \sqrt{\lambda_i}$ for $1 \leq i \leq n$.

The Eigenvectors of the matrix $M$ can provide a very useful decomposition of $D$. The Singular Value Decomposition of the matrix $D$ is of the form:

$$\boldsymbol{D = USV^T} \tag{3-11}$$

and $U$ and $V$ are orthogonal $m \times m$ and $n \times n$ matrices respectively, such that

$$\boldsymbol{U} = \{\boldsymbol{u_1}, \boldsymbol{u_2}, \dots, \boldsymbol{u_m}\} \text{ and } \boldsymbol{V} = \{\boldsymbol{v_1}, \boldsymbol{v_2}, \dots, \boldsymbol{v_n}\}$$

where $u_i = \frac{1}{\sigma_i} D v_i$ for all $i$ and $\sigma_i \neq 0$, and $v_i$ is the unit eigenvector where the eigenvector is defined as $Mv = \lambda v$. While $S$ is a $m \times n$ rectangular matrix with non-negative real numbers on the diagonal and zero elsewhere. The diagonal entries $S_{i,i}$ denoted by $\sigma_i$ are the non-zero singular values of $D$ arranged in decreasing order.

The condition number of the matrix $D$ can be defined as:

$$\boldsymbol{Condition(D)} = \frac{\boldsymbol{\sigma_{max}}}{\boldsymbol{\sigma_{min}}} \tag{3-12}$$

where $\sigma_{max}$ and $\sigma_{min}$ represent the maximum and minimum singular values of $D$, a high-condition number points to an ill-conditioned matrix whereas a low-condition number points to a well-conditioned matrix (Chen & Dongarra, 2005). This condition of the CS matrix will be used as a measure of the strength of the RIP property.

The following four steps illustrate a procedure to calculate the condition number, and implicitly compute the SVD, of any matrix:

**Procedure** (Computing the condition number of a matrix $D$)

1. Compute $M = D^T D$ for a matrix $D$.
2. Compute the eigenvalues from $\det(M - \lambda I) = 0$ and the eigenvectors $Mv = \lambda_i v$.
3. Find the singular values of $M$ by $\sigma_i = \sqrt{\lambda_i}$ and sort them by decreasing order.
4. Calculate the ratio of the maximum and minimum singular values of the matrix.

The statistical version of the Restricted Isometry Property (STRIP), which was introduced by Gan *et al.* (Gan, et al., 2009), performance bound in terms of the mutual coherence $\mu$ of the dictionary, which is an indicator of the dependence between columns of the matrix.

### 3.2.4 Coherence

Another numerical parameter that relates to compliance with CS requirements is the ***coherence*** of a matrix $D \epsilon \mathbb{R}^{m \times n}$ which provides information about the likelihood of guaranteed recovery of the sparse solution, and is defined as the largest absolute normalized inner product of pairs of columns $d_i$ and $d_j$, i.e.

$$\boldsymbol{\mu(D)} = \boldsymbol{max_{1 \le i < j \le n}} \frac{|\langle \boldsymbol{d_i}, \boldsymbol{d_j} \rangle|}{\|\boldsymbol{d_i}\|_2 \|\boldsymbol{d_j}\|_2} \tag{3-13}$$

The coherence of a matrix $D$ of size $m \times n$ is in the range $\mu(D) \epsilon [\sqrt{\frac{\boldsymbol{n-m}}{\boldsymbol{m(n-1)}}}, 1]$ where the lower bound is known as the Welch bound (Kutyniok, 2013). However, when $n \gg m$, the coher-

ence value has been shown ((Welch, 1974) and (Duarte & Eldar, 2011)) to be bounded below by $\mu(D) \geq \frac{1}{\sqrt{m}}$.

In order to be able to provide the relation between the Mutual Coherence and Restricted Isometry Property we need first introduce the following Lemma that will be used to prove this relationship.

**Lemma** 3.3: Let $\boldsymbol{x} = (\boldsymbol{x_1}, \boldsymbol{x_2}, \dots, \boldsymbol{x_n})$ be a vector of length $n$ then

$$\left[(\sum_{j=1}^{n}|\boldsymbol{x_j}|)^2 - \sum_{i=1}^{n}|\boldsymbol{x_i}|^2\right] = \boldsymbol{2}\sum_{i=1}^{n-1}\sum_{j>i}^{n}|\boldsymbol{x_i x_j}|$$

***Proof.***

$$\begin{aligned}\left[(\sum_{j=1}^{n}|\boldsymbol{x_j}|)^2 - \sum_{i=1}^{n}|\boldsymbol{x_i}|^2\right] &= [(|\boldsymbol{x_1}| + |\boldsymbol{x_2}| + \cdots + |\boldsymbol{x_n}|)^2 - (|\boldsymbol{x_1}|^2 + |\boldsymbol{x_2}|^2 + .. + |\boldsymbol{x_n}|^2)] \\ &= [|\boldsymbol{x_1}|^2 + |\boldsymbol{x_2}|^2 + \cdots + |\boldsymbol{x_n}|^2 + \boldsymbol{2}|\boldsymbol{x_1}||\boldsymbol{x_2}| + \boldsymbol{2}|\boldsymbol{x_1}||\boldsymbol{x_3}| + \cdots + \\ &\quad \boldsymbol{2}|\boldsymbol{x_{n-1}}||\boldsymbol{x_n}| - (|\boldsymbol{x_1}|^2 + |\boldsymbol{x_2}|^2 + \cdots + |\boldsymbol{x_n}|^2)] \\ &= \boldsymbol{2}|\boldsymbol{x_1}||\boldsymbol{x_2}| + \boldsymbol{2}|\boldsymbol{x_1}||\boldsymbol{x_3}| + \cdots + \boldsymbol{2}|\boldsymbol{x_{n-1}}||\boldsymbol{x_n}| \\ &= \boldsymbol{2}\sum_{i=1}^{n-1}\sum_{j>i}^{n}|\boldsymbol{x_i x_j}|\end{aligned}$$

**Theorem** 3.4 (Li, et al., 2012): If the dictionary $D \epsilon \mathbb{R}^{m \times n}$ has unit-norm columns and coherence $\mu = \mu(D)$, then $D$ satisfies the RIP of order $k$ with $\delta_k \leq (k-1)\mu$, for all $k < 1/\mu + 1$.

***Proof.*** The proof of this Theorem with details have not mentioned by (Li, et al., 2012) or other references. Therefore, we will prove the Theorem by assume the matrix $D$ of size $m \times n$ has $\mu$-coherence. The definition of RIP (3-10) is equivalent to the following inequality:

$$\left| \|\boldsymbol{Dx}\|_2^2 - \|\boldsymbol{x}\|_2^2 \right| \leq \boldsymbol{\delta_k}\|\boldsymbol{x}\|_2^2$$

For any $k$-sparse vector $x$ of length n,

$$\left| \|\boldsymbol{Dx}\|_2^2 - \|\boldsymbol{x}\|_2^2 \right| = \left| < \boldsymbol{Dx}, \boldsymbol{Dx} > - \|\boldsymbol{x}\|_2^2 \right|$$

The dot product $\boldsymbol{Dx}$ for the matrix $\boldsymbol{D}$ with columns $\boldsymbol{D_1}, \boldsymbol{D_2}, \boldsymbol{D_3}, \dots, \boldsymbol{D_n}$ is equal $\boldsymbol{Dx} = \boldsymbol{x_1 D_1} + \boldsymbol{x_2 D_2} + \cdots + \boldsymbol{x_n D_n}$. Hence,

$$\left| \|\boldsymbol{Dx}\|_2^2 - \|\boldsymbol{x}\|_2^2 \right| = \left| < \sum_{i=1}^{n} \boldsymbol{x_i D_i}, \sum_{j=1}^{n} \boldsymbol{x_j D_j} > - \|\boldsymbol{x}\|_2^2 \right|$$

Now, let

$< \sum_{i=1}^{n} \boldsymbol{x_i D_i}, \sum_{j=1}^{n} \boldsymbol{x_j D_j} > = < \boldsymbol{x_1 D_1 + x_2 D_2, \dots, + x_n D_n},\ \boldsymbol{x_1 D_1 + x_2 D_2, \dots, + x_n D_n} >$ and

Based on the properties of the inner product, we obtain that:

$$< \sum_{i=1}^{n} \boldsymbol{x_i D_i}, \sum_{j=1}^{n} \boldsymbol{x_j D_j} >= |\boldsymbol{x_1}|^2 < \boldsymbol{D_1, D_1} > + |\boldsymbol{x_2}|^2 < \boldsymbol{D_2, D_2} > + .. + |\boldsymbol{x_n}|^2 < \boldsymbol{D_n, D_n} >$$

$$+ 2 \sum_{i=1}^{n-1} \sum_{j>i}^{n} |\boldsymbol{x_i x_j}| < \boldsymbol{D_i, D_j} >$$

$$= \sum_{i=1}^{n} |\boldsymbol{x_i}|^2 < \boldsymbol{D_i, D_i} > + 2 \sum_{i=1}^{n-1} \sum_{j>i}^{n} |\boldsymbol{x_i x_j}| < \boldsymbol{D_i, D_j} >$$

Therefore,

$$\left| \|\boldsymbol{Dx}\|_2^2 - \|\boldsymbol{x}\|_2^2 \right| = \left| \|\boldsymbol{x}\|_2^2 < \boldsymbol{D_i, D_i} > + 2 \sum_{i=1}^{n-1} \sum_{j>i}^{n} |\boldsymbol{x_i x_j}| < \boldsymbol{D_i, D_j} > - \|\boldsymbol{x}\|_2^2 \right|$$

Since $\|\boldsymbol{D_i}\|_2 = \boldsymbol{1}$ then $< \boldsymbol{D_i, D_i} > = \boldsymbol{1}$ and that lead to:

$$\left| \|\boldsymbol{Dx}\|_2^2 - \|\boldsymbol{x}\|_2^2 \right| = \left| \boldsymbol{2} \sum_{i=1}^{n-1} \sum_{j>i}^{n} |\boldsymbol{x_i x_j}| < \boldsymbol{D_i, D_j} > \right|$$

Since, the matrix *D* has $\mu$-coherence and $|< \boldsymbol{D_i, D_j} >| \leq \boldsymbol{\mu}$ then we get that:

$\left| \|\boldsymbol{Dx}\|_2^2 - \|\boldsymbol{x}\|_2^2 \right| \leq \boldsymbol{2\mu} \sum_{i=1}^{n-1} \sum_{j>i}^{n} |\boldsymbol{x_i x_j}|$, then by Lemma 3.3, we get that

$$\leq \boldsymbol{\mu}[(\sum |\boldsymbol{x_j}|)^2 - \|\boldsymbol{x}\|_2^2]$$

$$\leq \boldsymbol{\mu}[\boldsymbol{k}\|\boldsymbol{x}\|_2^2 - \|\boldsymbol{x}\|_2^2] \qquad \text{Since } \sum |\boldsymbol{x_j}| = \|\boldsymbol{x}\|_1 \leq \sqrt{\boldsymbol{k}}\|\boldsymbol{x}\|_2$$

$$\leq \boldsymbol{\mu(k-1)}\|\boldsymbol{x}\|_2^2$$

Therefore, for any *k*-sparse vector $\boldsymbol{x}$ the RIP inequality satisfy the definition of Restricted Isometry Constant (RIC) by $\boldsymbol{\delta_k = \mu(k-1)}$**,** because $\boldsymbol{\delta_k}$ is the smallest constant.

■

In practice, statistical $\mu$-coherence for a hundred randomly selected sub-matrices from the overcomplete CS dictionary will be used to ensure sparse recovery. Coherence, Condition number and Spark, will be our preferred measures to estimate the strength of RIP of the various dictionaries in the thesis.

## 3.3 A Selection of RIP Dictionaries

Compressive sensing-compliant dictionaries are pre-dominantly generated using certain known random distributions, which are among the most dominant in the applications. However, there are many deterministic RIP dictionaries that have been proposed in the literature

such as deterministic binary dictionaries via algebraic curves over finite fields (Li, et al., 2012), and partial bounded orthogonal matrices in which the sensing operators are obtained by choosing $M$ rows uniformly at random from a normalized $N \times N$ Fourier transform matrices (Rauhut, 2010).

In spite of, the fact that random matrices are important in many applications, randomness has drawbacks. First of all these matrices are only known to satisfy the RIP property of order $k$, where $k \leq Cm\ log(n/m)$ and $C$ is a constant, but with high probability (Li, et al., 2012). Moreover, there is no efficient algorithm testing the RIP. These provide the incentive to construct new overcomplete dictionaries that satisfy the property of CS by design and avoid these difficulties.

Now, we briefly describe well-known different random methods for constructing dictionaries that are known to be CS-compliant in terms of RIP. These include Gaussian, Toeplitz and Circulant Random Matrices. The suitability of these dictionaries for the objectives of this thesis will be investigated in the remaining chapters of the thesis. In particular, these dictionaries are used to reconstruct super resolution images from low-resolution face images, and we shall test the consequences for face recognition in uncontrolled conditions. In Chapter 4, we shall describe other dictionaries, which will be introduced, and their RIP credentials in terms of the above indicators of coherence and condition numbers will be investigated.

### 3.3.1 Gaussian Random Matrix

A widely used RIP dictionary is the Gaussian Random Matrix (GRM). For Gaussian Random Matrix with normalized columns $\|x_l\|_2 = 1$, the entries $x_{i,j}$ of the size $m \times n$ matrix ($m \ll n$) are independently sampled from a normal distribution $x_{i,j} \sim N(0,1/m)$. The probability density function of the normal distribution with mean $\bar{x}$ and variance $\boldsymbol{\sigma^2}$ is defined as follows:-

$$\mathbf{f}(\boldsymbol{x}) = \frac{\mathbf{1}}{\sqrt{\mathbf{2\pi\sigma^2}}}\mathbf{e}^{\frac{-\mathbf{1}}{\mathbf{2}}\left(\frac{\boldsymbol{x}-\bar{\boldsymbol{x}}}{\boldsymbol{\sigma}}\right)^{\mathbf{2}}}$$

GRM satisfies the sufficient condition for CS property with a (near) optimal order $k$ and therefore allows sparse recovery using $l_1$-minimization. In the following chapters, we shall implement a procedure recover a super resolved image from an image with lower resolution via sparse representation by using two overcomplete Gaussian dictionaries of different size, both constructed from a zero mean Gaussian distribution with variance $1/m$.

### 3.3.2 Toeplitz - Circulant Random Matrix

The Gaussian matrix is known to provide an optimal condition for the minimal number of required samples for sparse recovery (Rauhut, 2010). However, some CS applications often do not allow the use of "completely" random matrices, but impose certain constraints on the adopted measurements and limit the amount of randomness that can be used such as the entries of the first row of the matrix are only random. Practically this points to the use of a structured random measurement matrix. In (Bajwa, et al., 2007) they have shown that the use of under-determined Toeplitz-structured matrices for acquiring signals is sufficient to recover under-sampled sparse signals. These matrices satisfy the RIP condition but with high probability and have been proposed for certain applications such as wireless communications and radar (Rauhut, 2010), (Yu, et al., 2010). Toeplitz and Circulant matrices of the size $k \times n$ are of the form, respectively:

$$\mathrm{T} = \begin{bmatrix} t_n & t_{n-1} & \cdots & t_1 \\ t_{n+1} & t_n & \cdots & t_2 \\ & \vdots & & \\ t_{n+k-1} & t_{n+k-2} & \cdots & t_k \end{bmatrix}, \text{ and } \mathrm{C} = \begin{bmatrix} t_n & t_{n-1} & \cdots & t_1 \\ t_1 & t_n & \cdots & t_2 \\ & \vdots & & \\ t_{n-1} & t_{n-2} & \cdots & t_k \end{bmatrix}$$

In both types, every left-to-right descending diagonal is a constant and the Toeplitz matrix is a special case of a Circulant matrix. In fact, Toeplitz matrices can be embedded in a Circulant matrix, as illustrated below

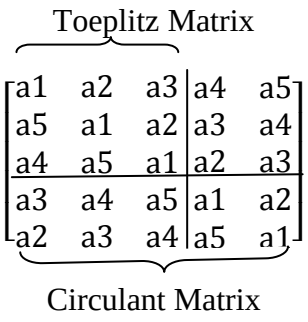



**Figure 3.3:** Example of Circular Matrix of size $\mathbf{5 \times 5}$ with $\{\mathbf{a_i}\}_{\mathbf{i=1}}^{\mathbf{5}}$ numbers.

For image reconstruction, the overcomplete dictionaries are generated as TCRM matrices, by selecting the first row using the standard Gaussian distribution with zero mean and variance one. The rest of the rows are permuted versions of it as required above.

## 3.4 Summary and Conclusion

Compressive sensing is a new paradigm in image acquisition/processing with specific implications for resolution enhancement and aims to reconstruct a sparse signal from a few non-adaptive linear measurements well below Shannon-Nyquist requirements. Basic mathematical concepts of CS theory have been studied in this chapter with focus on the fundamental properties that ensure unique sparse recovery for underdetermined matrices. New proof to the Theorem (3.4) that describes the relation between coherence and sufficient CS property has been constructed. In addition, the relationship between NSP property and linear independence vectors in a dictionary has been illustrated and proved. Finally, we have described examples of overcomplete random dictionaries that satisfy the RIP property and are used to reconstruct super resolve images in the following chapters.

# Chapter 4

## IMAGE QUALITY/RESOLUTION ENHANCEMENT USING CS-BASED SUPER RESOLUTION

Image super resolution (SR) techniques play an important role in many applications, such as medical imaging, surveillance systems and particularly in face recognition where resolution is one of the most influential factors affecting face recognition at a distance. As described earlier, the term "super resolution" refers to the process of producing a high-resolution (HR) image from a single or a collection of low-resolution (LR) images. The performance of face recognition systems depends, among many other factors, on the quality (i.e., discriminating ability) of a set features extracted from the face image, and not necessarily on the quality of the image itself. Therefore, it makes sense to investigate the use of SR to enhance face image features with an ultimate aim of increasing the accuracy of face recognition at a distance. First, we will focus on the use of super resolution techniques to enhance face image quality with respect to image resolution and degradation.

Over the past few years, several image-resolution enhancement techniques with different degrees of success, depending upon application requirements, have been proposed in the literature. Examples of such techniques include the traditional non adaptive interpolation techniques (e.g. bilinear, bi-cubic and nearest interpolation) where the unknown pixel values are estimated based on their neighbouring pixels (Levy, 2010), iterative methods (e.g. back-projection iterative interpolation SR method) (Bannore, 2009), (Irani & Peleg, 1991), and SR based on sparse representation from compressive sensing (CS) dictionaries (Yang, et al., 2010), (Yang, et al., 2008), (Zhang, et al., 2012), (Studer, 2010). In this chapter, the *CS Dictionaries* based SR technique will be studied with the aim of reconstructing a higher resolution version of a low-resolution image.

Yang et al. (Yang, et al., 2010) and (Yang, et al., 2008), used domain-relevant HR images to create dictionaries to super-resolve LR images. The approach of using training image data to

create CS dictionaries is not convenient or scalable since one has to find application-relevant training images. The aim of this chapter is to question the need for image-learnt dictionaries for resolution enhancement tasks. In other words, this chapter will investigate the possibility of creating compressive sensing dictionaries that do not depend on training images. We propose a number of approaches to create dictionaries that satisfy CS property and are independent of training data. These proposed approaches will be used for two purposes: 1) to enhance quality and resolution of low-resolution images; and 2) to overcome the problem of differences in resolution between probe and gallery images that are typically encountered in face recognition in uncontrolled conditions. In the second scenario, we will investigate the super resolving of images as well as face feature vectors obtained from images for face recognition at a distance. The following summarises the main contributions of this chapter, and in turn, of the thesis.

*First*: we introduce the concept of constructing full spark non-adaptive dictionaries as well as random orthonormal block matrices, i.e. without using image information, that satisfy by design the Restricted Isometry Property (RIP) sufficient for recovery of compressively sensed images.

*Second*: we use existing dictionaries such as Toeplitz Circular Random matrices as well as random dictionaries generated by Gaussian random numbers for image resolution enhancement. These dictionaries have been previously used in CS to recover a unique k-sparse vector (see Chapter 3, Section 3.3), but not for image resolution enhancement. We will compare the performance of the proposed dictionaries with those that are learnt from face image databases, using elaborate procedures, in terms of the quality of their super-resolved images. Moreover, the "strength" of meeting the required CS properties (RIP) for the proposed dictionaries will be illustrated.

*Third*: we construct overcomplete dictionaries on the high/low frequency domain of the training images to enhance the high/low frequency components of the input LR in terms of image quality. Although the CS dictionaries created using images in the spatial domain produce slightly better quality super resolved images than the proposed dictionaries, we will investigate, in the next chapter, the effectiveness of the proposed CS dictionaries in recovering high frequency feature vectors of LR images for face recognition at a distance as well as synthetically degraded LR images using common models of image degradation.

We shall demonstrate that the proposed CS dictionaries constructed without using training images as well as the variety of existing random dictionaries are able to improve image quality as much as, if not better than the learning dictionaries. Hence, we will argue that there is no need to use training images to construct dictionaries. We will show that the non-CS based iterative SR method in the spatial domain is able to reconstruct image with good quality from multiple LR images. Furthermore, we evaluate the use of iterative SR method in the wavelet domain, but this approach does not always produce good quality images compared to its counterpart in the spatial domain. Therefore, in this thesis, the performance of the non-CS based iterative SR method in the spatial domain as well as the standard interpolation method will be used to compare the effectiveness of the proposed dictionaries. We will demonstrate that the various types of CS dictionaries based SR technique are able to improve image quality and image resolution better than the iterative SR method in the spatial domain and even better than bi-cubic interpolation method.

The rest of this chapter is organised as follows. We begin with an introduction to image super resolution process. Then in section 4.2, we describe the proposed iterative SR (IISR) method in the wavelet domain and test its ability to reconstruct images with good quality. Following which, we shall explain the use of SR by learning dictionary in spatial and wavelet domains to enhance face image quality. The proposed approaches to construct dictionaries that do not depend on training images are introduced in section 4.5. In section 4.6, we will briefly discuss and compare the various proposed dictionaries and the learning dictionaries in terms of their strength in meeting the CS properties. Finally, the results of experiments conducted on two databases to evaluate the efficiency of the proposed dictionaries to enhance face image quality will be presented in section 4.7.

## 4.1 Image Super Resolution

Super resolution is an inverse problem for obtaining a HR image from a degraded LR image(s). SR enhancement techniques are used as a pre-processing step to recover lost high-frequency components of LR images in such a way that the result approximates the original image as closely as possible. Generally, SR techniques used to obtain a HR image from an observed LR input image $\boldsymbol{y}$ can be modelled as a solution $\boldsymbol{x}$ of the matrix equation:

$$\boldsymbol{y} = \boldsymbol{SBx} + \boldsymbol{\eta} \tag{4-1}$$

where $B$ is the point-spread function with a blurring effect, $S$ is the down-sampling function and $\eta$ is the additive noise. The main challenge in recovering $x$ is the modelling of the unknown blurring function $B$.

Gaussian low-pass function is a well-known frequency-domain smoothing filter that has a blurring effect on images (Gonzalez & Woods, 2002) and is defined as follows:

$$\boldsymbol{G(u,v) = e^{-D^2(u,v)/2\sigma^2}} \tag{4-2}$$

where $\boldsymbol{D^2(u,v) = (u^2 + v^2)}$ and $\boldsymbol{\sigma}$ is the standard deviation of the Gaussian curve. Gaussian function can be considered a suitable model of blurring/degradation to use in SR procedures, but this function has a low level of blurring effect on images and does not reflect the severe degradation conditions seen in surveillance scenarios. Another suitable model of degradation can be based on the use of atmospheric turbulence function of different strengths, which result in more severe degradation effect than the Gaussian functions. Turbulence function is a degradation function which models the environmental conditions caused by variations in temperature, wind speed and exposure time. In the frequency domain, the turbulence functions (Gonzalez & Woods, 2002) are of the form:

$$\boldsymbol{H(u,v) = e^{-k_1(u^2+v^2)^{5/6}}} \tag{4-3}$$

where $k_1$ is a constant that reflects the severity of blurring. We label degradation based on the experimental results as *severe* if $k_1 \in (0.04, 0.09]$, *medium* (similar to most Gaussian blurring functions) if $k_1 \in (0.02, 0.04]$ and minor if $k_1 \in (0, 0.02]$. The small images in Figure 4.19 are the down sampled degraded images after applying $H$ for different values of $k_1$. In all the experiments, we shall adopt two models of degradation; the first is a Gaussian filter of size seven and the second is atmospheric turbulence with blurring of different extents. SR algorithms aim to maximize peak signal to noise ratio (PSNR) or minimize the mean squared error between the hypothetical HR image and the reconstructed SR image. Therefore, this chapter focuses on the performance of SR methods in terms of image quality, and later in the following chapter for face recognition under uncontrolled conditions.

## 4.2 Iterative Super Resolution Method in the Wavelet Domain

The performance of face recognition schemes are adversely affected by image degradation such as blurring of the edges that represent the most discriminating facial features. Therefore,

in this section, the iterative SR technique, described in Chapter 2, will be used in the wavelet domain to address the problem of image resolution and to reconstruct a HR image from the wavelet subbands of two degraded LR images. The two distorted LR images ($LR_1$ and $LR_2$) of an individual are obtained from the original image by applying the degradation function $H$ in equation (4-3) with different $k_1$-values and then down sampling the image by taking even and odd indexed pixels.

The following steps explain the IISR method in the wavelet domain.

1. Transform the two LR images into wavelet domain by decomposing each image into four subbands of different frequency ranges (approximate and detailed subbands) using discrete wavelet transform (DWT).
2. Combine the low subband (LL) of each of the two images to generate a LL subband of large size than the subband of LR image by:
    a. Place the coefficients of each LL subbands into a HR grid (i.e. fusion subband) based on the information that transfer the image into LR images, (see Chapter 2).
    b. Use bi-cubic interpolation to complete the HR grid by interpolating the missing pixels (unknown coefficients). Note that the size of the combined subband depends on the size of the wavelet subband of the LR images, wavelet filter, level of decompositions, as well as depends on the size of the enlarge input images that we want it.
3. Perform step 2 for the three non-LL subbands.
4. Apply the inverse discrete wavelet transform (IDWT) to generate initial HR image.
5. Use the back project iterative method to remove possible artifacts and/or blurring degradation of the inversed image and obtain an image with higher resolution.

Different wavelet filters of the same family (daubechies-2 (db2), daubechies-4 (db4), daubechies-6 (db6), and daubechies-8 (db8)) have been tested at wavelet decomposition levels from 1 to 3. The choice of wavelet filter has been considered based on the wavelet filter length to test the effect of such on the used images. In order to, objectively evaluate the performance of the iterative SR algorithm in the spatial and wavelet domain; a number of experiments were conducted on the test images shown below in Figure 4.1.

The well-known bi-cubic interpolation method was used as a baseline to enhance the resolution of the input LR images. The reconstruction reliability is quantified using the well-known quality measure PSNR, whilst the original test images are used as ground truth to calculate

PSNR values. The average PSNR values obtained by each of the SR methods are presented in Figure 4.2.

**Figure 4.1:** Original Test Images

It can be seen from Figure 4.2 that the IISR methods in the spatial and wavelet domain produce superior quality improvement when compared to the standard bi-cubic interpolation technique. However, the IISR method in the spatial domain slightly outperforms the IISR in the wavelet domain for each wavelet decomposition levels and for each level of degradation function.

The various charts in Figure 4.2 show the quality of the super resolved images by IISR methods and bi-cubic technique decreases with increased level of blurring. Furthermore, with mild/medium and severe degradation, there is a small difference in terms of reconstructed image quality between different wavelet filters for each levels of wavelet decomposition.

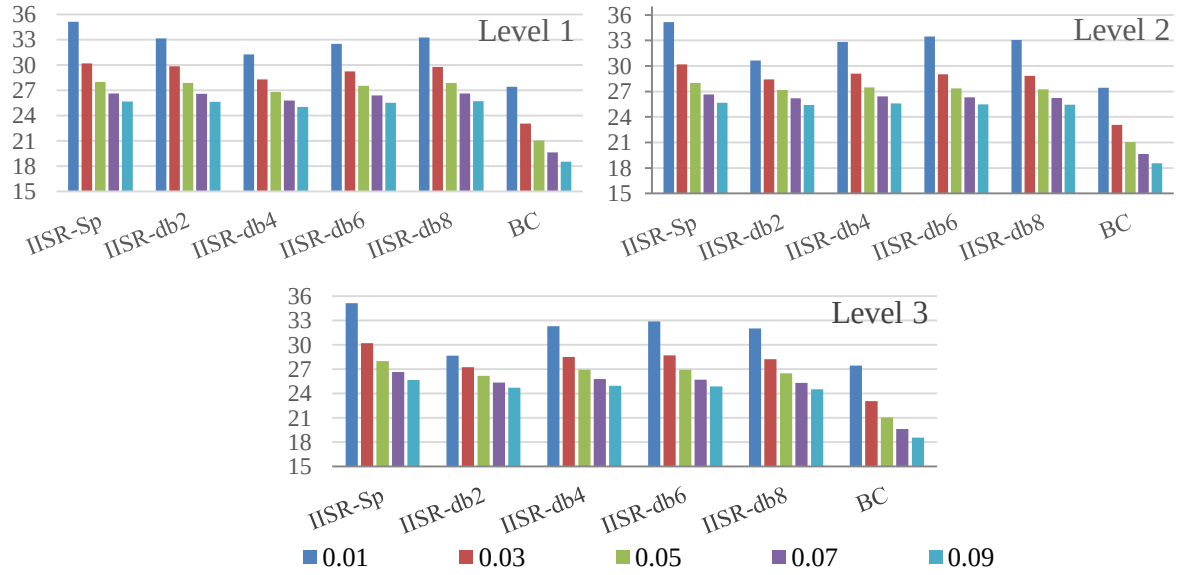



**Figure 4.2:** Average PSNR values for the super resolved images from two LR images with different $k_1$-Turbulence blurring and with different wavelet levels.

Super resolved images for two different subjects are presented in Figure 4.3 and Figure 4.4. The results of other test images (in Figure 4.1) and other levels of degradation are similar to these images and therefore are not shown here. Hence, we shall only demonstrate the outcome of the experiments on these images where the LR images have low levels of degradation.

From the results in Figures (4.3 and 4.4), we can observe that the IISR algorithm in the spatial domain produced sharper results and demonstrated less visual blurring than the image interpolation. Furthermore, IISR in the spatial domain provides more image details than the one enhanced in the wavelet domain. The super resolved images by wavelet domain algorithm with different filters produce images with number of artifacts or geometric degradation. However, the reconstructed image by Haar filter (db2) in level one wavelet decomposition is slightly better than one reconstructed by other wavelet filters. Since the size of the filter has an effect on the image, where the Haar filter of length 2 is able to capture sharp changes (high frequency) in the LR images better than the others filters.

In general, the experimental results show the effectiveness of IISR techniques in the spatial domain and its ability to achieve better results in terms of both measured quality and visual perception. Furthermore, there is no significant difference in terms of the quality of super-resolved images based on the 4-wavelet filters at different decomposition levels. The interpolation method yielded the worst performance overall. Hence, the IISR method in the spatial domain will be used to super-resolve images in all the experiments in this chapter.

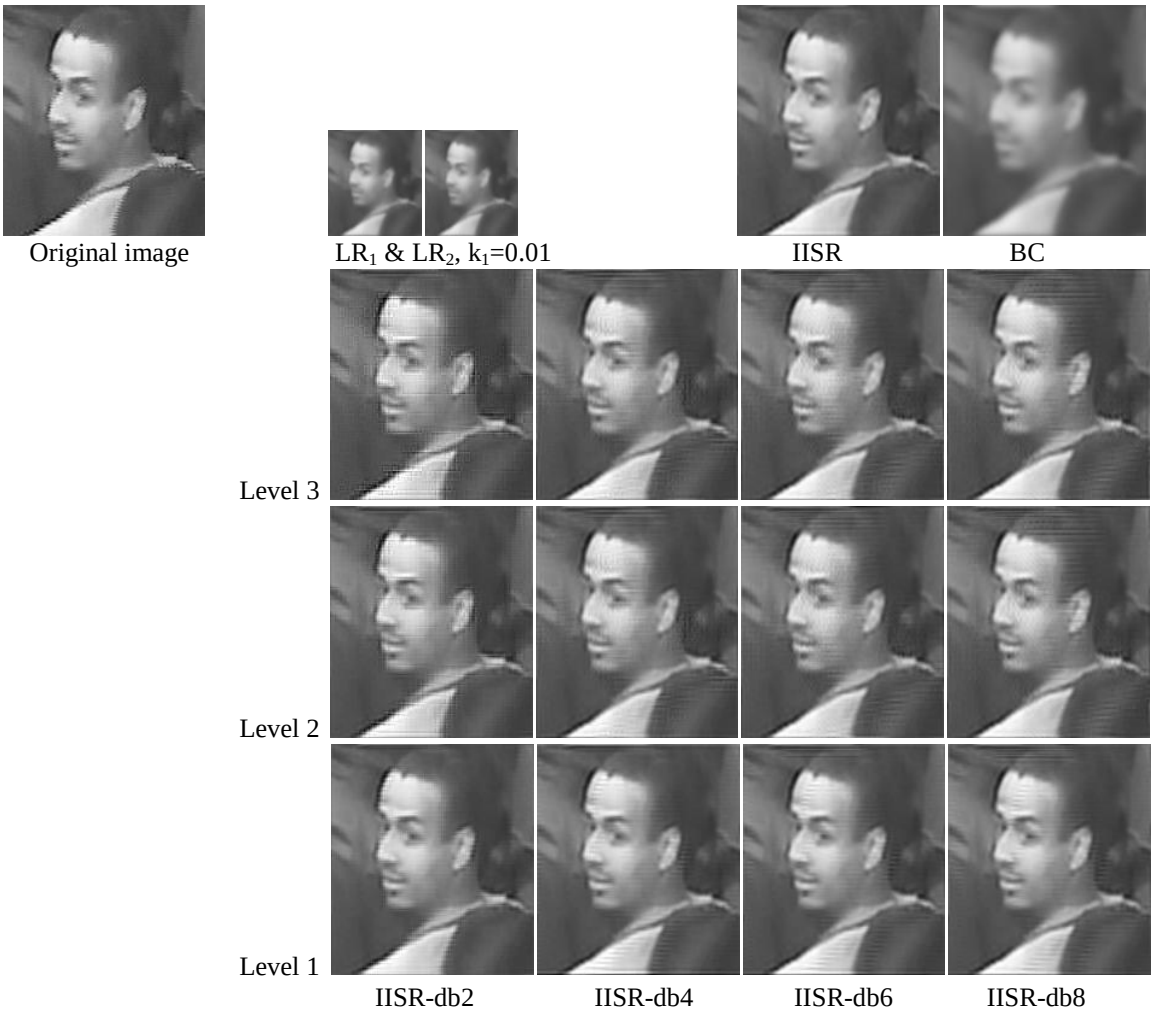



**Figure 4.3:** Example 1, comparison between super resolved image by interpolation method & IISR approaches in the spatial and wavelet domain from LR images with low-level of degradation.

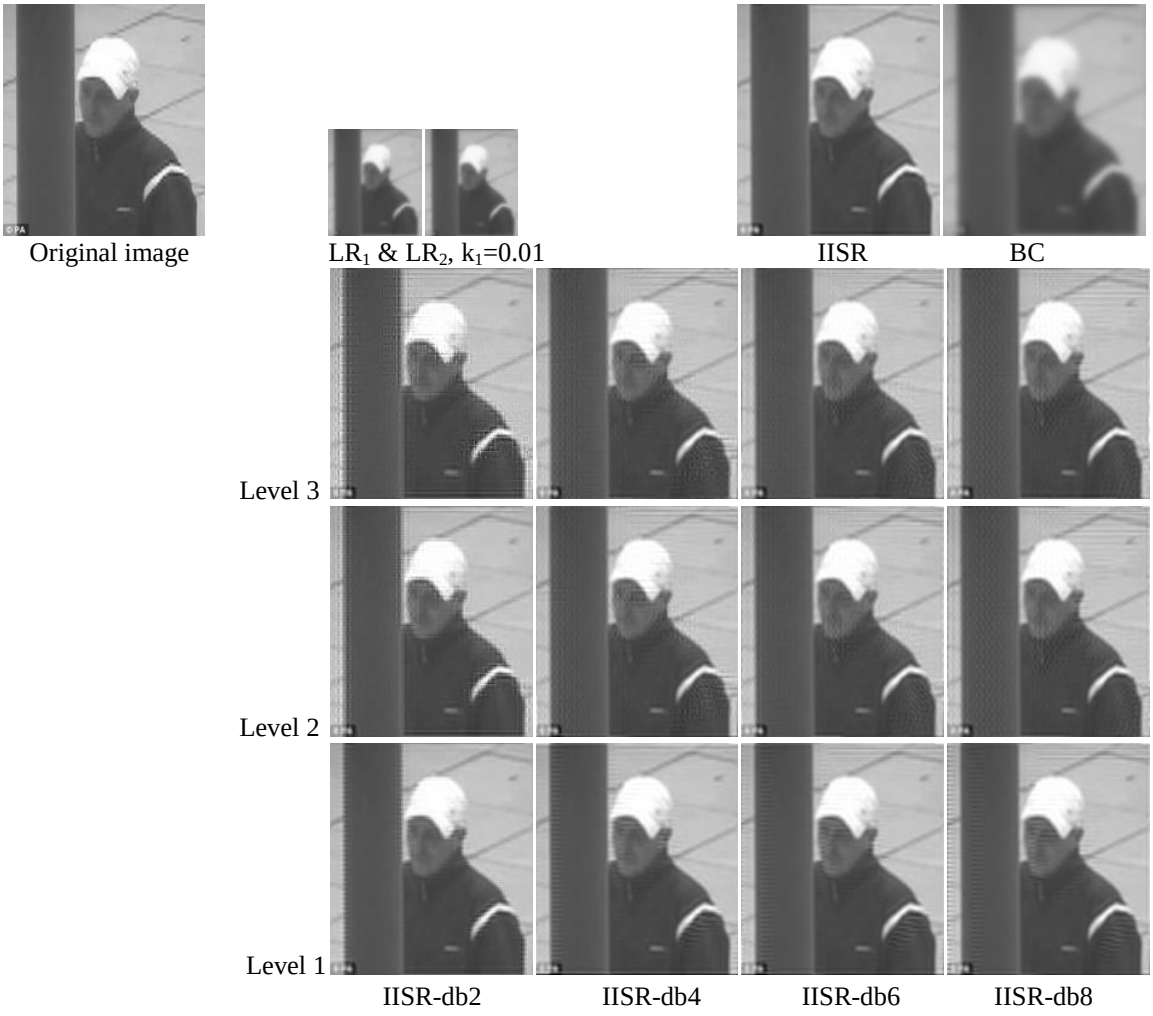



**Figure 4.4:** Example 2, comparison between super resolved image by interpolation method & IISR approaches in the spatial and wavelet domain from LR images with a low-level of degradation.

## 4.3 Super Resolution by Learning Dictionary in the Spatial Domain

Influenced by the latest advances in sparse representation theory and algorithms, the Learning Dictionary approach is an underdetermined matrix design process that uses training image patches to create dictionaries. Yang et al. in (Yang, et al., 2010), and (Yang, et al., 2008) prepared a pair of learning dictionaries to reconstruct a SR image from a single LR input image, based on sparse signal representation; we will refer to this method as LD-Sp. The learning dictionaries based on SR method has two main stages; the first stage prepares the two dictionaries $D_H$ and $D_L$ and the second stage super resolves a single input LR image to produce a HR image using the two dictionaries. Figure 4.6 illustrates the two stages of SR by LD-Sp algorithm, which we will explain in detail.

**Stage 1: Prepare two Overcomplete Dictionaries**

The compact dictionaries $D_H$ and $D_L$ in general can be generated from large matrices of the form $X_H = \{x_1, x_2, \dots, x_n\}$ and $Y_L = \{y_1, y_2, \dots, y_n\}$ of sampled HR image patches $x_i$ and their corresponding LR image feature patches $y_i$ by the following procedure:

1. Build $X_H$ for sufficiently large number of random raw patches of size $5 \times 5$ and turn each patch into a column vector. The patches (the patches overlap) are carefully selected from a set of HR training images of the same statistical nature to the input LR image (in this case face images).
2. Generate a LR image set by:
   2.1 blurring and down sampling the HR training images, and
   2.2 using standard interpolation technique to enlarge the down-sampled images.
3. Apply feature extraction procedure, which consists of four 1D filters chosen as first and second order gradient filters, to the up-sampled version of the LR images to generate four feature images for each image.
4. Generate $Y_L$ by the randomly selected patches from the feature images that extracted from the up-sampled LR images, and subtract the mean pixel value from each patch to ensure that the dictionary represents image textures. Then, turn the four feature image patches into a single vector by concatenation to arrive at the final representation of the LR patch.
5. Generate th**e** large size matrix $\mathrm{X} = \begin{bmatrix} \frac{1}{\sqrt{\mathrm{N}}}\mathrm{X_H} \\ \frac{1}{\sqrt{\mathrm{M}}}\mathrm{Y_L} \end{bmatrix}$ where $M$ and $N$ are the dimensions of the HR and LR image patches in vector form, and
6. Use the learning strategy of the single dictionary (Yang, et al., 2010), (Lee, et al., 2007) shown in Figure 4.5, which is an iterative algorithm starting with an initial dictionary, to generate and fix the size S of the two dictionaries $D_H^{M\times S}$ and $D_L^{N\times S}$ from the large dictionary X .

1. Start with initial dictionary $D$ with Gaussian random matrix.
2. Fix $D$ and find $Z$, where $Z = min\|X - DZ\|_2^2 + \lambda\|Z\|_1$
3. Fix $Z$ and update $D = min\|X - DZ\|_2^2$
4. Iterate steps 2, 3 until converge.
5. Find the $D_H$ and $D_L$ from $D = \begin{bmatrix} \frac{1}{\sqrt{N}} D_H \\ \frac{1}{\sqrt{M}} D_L \end{bmatrix}$

**Figure 4.5:** Single Dictionary Strategy Algorithm

**Stage 2: Reconstruct a HR image**

To reconstruct the output HR image, the SR scheme works as follows: (1) Up-sample the input LR image using bi-cubic interpolation technique. (2) Apply the one dimension gradient filters to highlight the edges in different direction to the interpolated image and generate four feature images. (3) Starting from the upper left corner of the feature images, divided the images into overlap patches. Then (4) recover HR image patches by:

- For each LR image patch $y$ (or features extracted from it) finds a sparse representation $\alpha$ with respect to low-dictionary $D_L$ by $l_1$-minimization problem (i.e. by using linear programming algorithm). This algorithm start with an initial basis of linear independent vectors and iteratively improve by swapping one column in the initial matrix by one in the dictionary (see Chapter 3) since the solution to the linear programming is identical to finding this basis.
- The coefficients $\alpha$ of this representation with the corresponding patches in $D_H$ use to generate the output HR image patch $x = D_H\alpha$ as a linear combination of k-columns to the dictionary.

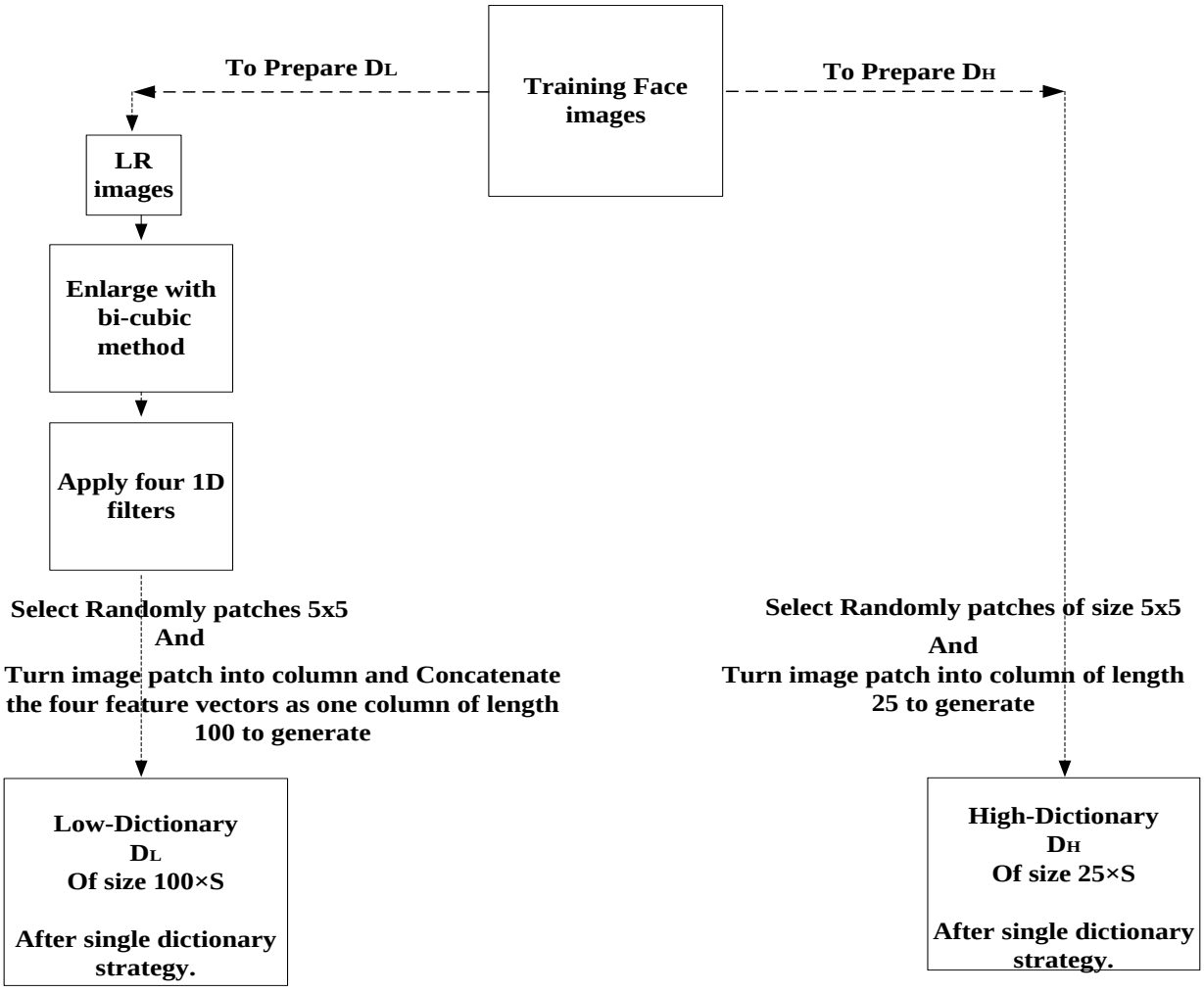
The computed sparse representation adaptively selects the most relevant patch bases in the dictionary to best represent each patch of the given LR image. For global reconstruction, restrict the use of the iterative back-project (IBP) method to a single image to remove possible artifacts from the local sparse and to eliminate the reconstruction errors in the HR image.

In general, the standard single LR image IBP scheme works as follows:

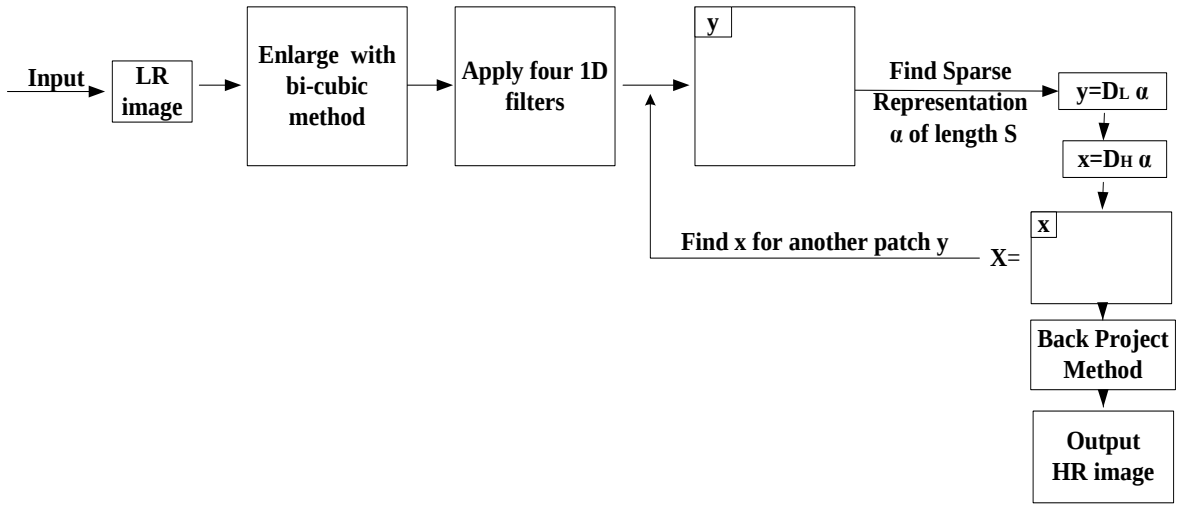
1. Generate the initial HR image $x_0$ (here the image obtained from the local sparse is the initial image), and at the $n^{th}$ iteration, $n > 0$
2. Convolve the image $x_{n-1}$ with an appropriate degradation function.
3. Down sample the resulting image to obtain $LR_{(n)}$, and

4. Obtain the error image $x_e$ of the same size of the initial image from up sampling the difference$\left(LR - LR_{(n)}\right)$.
5. Update the HR image by calculating $x_n = (x_{(n-1)} + x_e * p)$ where $p$ is the restoration filter which is used to remove blurring effects in the image $x_e$.
6. Finally, input the new HR image to the next iteration cycle.

The iteration procedure terminates either when the error term $\left(LR - LR_{(n)}\right)$ is less than/or equal a certain threshold or the number of iterations reached a fixed maximum number. In (Makwana & Mehta, 2013), interesting variants of the IBP have been proposed. They presented an approach based on combining the IBP method with the Canny edge detection and Gabor filter. At each iteration, the authors added additional terms representing high frequency information in $x_0$ by using Canny edge information and Gabor filter to the back-project error image for improve image quality.




**Stage 1: Prepare two dictionaries**




**Stage 2: Reconstruct HR image from the input LR image**

**Figure 4.6:** Illustrates the two stages of SR by learning dictionary method in the spatial domain.

### 4.3.1 Feature Representation for LR Image Patches

The high-frequency components of the LR image are the most important for predicting the lost high frequency content in the target HR images. The first and second-order gradients of the LR image have been used during sparse coding and in the process of creating the $D_L$ dictionary to highlight/model the edges in different directions (Chang, et al., 2004). The first order $\nabla_{13}$ and second order $\nabla_{13}^2$ gradient vector of the centre value $z_{13}$ in the $5 \times 5$ block (see Figure 4.7) can easily be derived as follows:

$$\nabla_{13} = \begin{bmatrix} (z_{14} - z_{13}) + (z_{13} - z_{12}) \\ (z_{18} - z_{13}) + (z_{13} - z_8) \end{bmatrix} = \begin{bmatrix} z_{14} - z_{12} \\ z_{18} - z_8 \end{bmatrix} = \begin{bmatrix} 0 & -1 & 0 \\ -1 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix}$$

$$\nabla_{13}^2 = \begin{bmatrix} (z_{15} - z_{13}) - (z_{13} - z_{11}) \\ (z_{23} - z_{13}) - (z_{13} - z_3) \end{bmatrix} = \begin{bmatrix} z_{15} - 2z_{13} + z_{11} \\ z_{23} - 2z_{13} + z_3 \end{bmatrix} = \begin{bmatrix} 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & -2 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \end{bmatrix}$$

Therefore, the four 1-dimension filters used to extract the derivatives are:

$$f_1 = [-1 \;\; 0 \;\; 1], f_2 = f_1^T = [-1 \;\; 0 \;\; 1]^T, f_3 = [1 \;\; 0 \;\; -2 \;\; 0 \;\; 1], f_4 = f_3^T = [1 \;\; 0 \;\; -2 \;\; 0 \;\; 1]^T.$$

where the superscript "*T*" means transpose. Applying these four filters on the image patches, we get four description feature vectors for each patch, which are concatenated into one vector as the final representation of the LR patch.

| | | | | |
|---|---|---|---|---|
| $z_1$ | $z_2$ | $z_3$ | $z_4$ | $z_5$ |
| $z_6$ | $z_7$ | $z_8$ | $z_9$ | $z_{10}$ |
| $z_{11}$ | $z_{12}$ | $\boldsymbol{z_{13}}$ | $z_{14}$ | $z_{15}$ |
| $z_{16}$ | $z_{17}$ | $z_{18}$ | $z_{19}$ | $z_{20}$ |
| $z_{21}$ | $z_{22}$ | $z_{23}$ | $z_{24}$ | $z_{25}$ |

**Figure 4.7:** A **5 × 5** local neighbourhood in the LR image for computing the first and second order gradients of the pixel at the centre value $Z_{13}$.

The following section gives a brief explanation of the SR method based dictionaries in the wavelet domain to super-resolve each wavelet subbands of a degraded LR image. A number of experiments on the test images in Figure 4.1 with different $k_1$of degradation have been conducted to assess the viability of this method in terms of image quality.

## 4.4 Super Resolution by Learning Dictionary in the Wavelet Domain

The fact that facial features used in recognition systems are associated with high/low frequency content suggest that building dictionaries in the LL and non-LL frequency wavelet subbands to enhanced subbands of hypothetical HR face images could result in an improved image quality. Therefore, this section investigates the use of SR based learning dictionaries in the wavelet domain, hereafter referred to as LD-WD, to reconstruct feature vectors (i.e. wavelet subbands) that are extracted from an input LR image. Here, the learning dictionaries will be building based on wavelet subbands obtained from a training set of HR face images. The inverse wavelet transformation will be used to reconstruct an image of higher resolution than the corresponding LR input image. In terms of face recognition, enhanced wavelet coefficients will be used in the following chapter for face identification at a distance.

In a similar manner to SR based on LD in the spatial domain, we can build two overcomplete dictionaries $D_H$ and $D_L$ for each wavelet subband of the training image sets. The high-dictionary $D_H$ will be generated using randomly selected patches from the corresponding subband of the training set of HR images. To create the corresponding $D_L$ for each subband, we synthetically generate the LR patches by blurring the HR training images with a Gaussian filter and down sampling them. The SR reconstruction process of a subband of a given LR face image is achieved by first dividing the subband into patches with an overlap of one pixel to avoid discontinuities during reconstruction. Secondly, a sparse representation of these patches is computed using $D_L$ and the corresponding high patch bases $D_H$ will be combined according to the sparse vector in order to generate the output subband. Finally, for global reconstruction, we used the back project iterative method after inverse DWT to eliminate the reconstruction errors in the HR image results. The sparse representation is obtained by minimizing:

$$\hat{\boldsymbol{\alpha}} = \min \boldsymbol{\lambda}\|\boldsymbol{\alpha}\|_1 + \tfrac{1}{2}\|\boldsymbol{F}\boldsymbol{D}_{\boldsymbol{L}}\boldsymbol{\alpha} - \boldsymbol{F}\boldsymbol{y}\|_2^2 \tag{4-4}$$

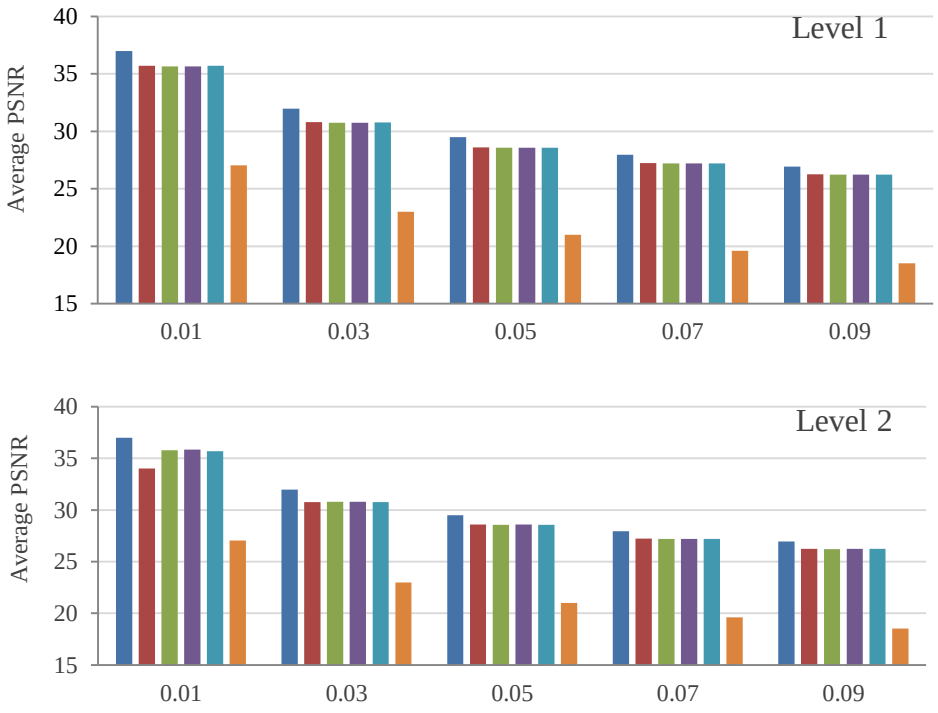
where the parameter $\lambda = 0.15$, is responsible for balancing the sparsity of the solution and the reliability of the approximation $y$, and $F$ is formed by the concatenation of four 1- dimension filters, while $\alpha$ is a sparse vector.

The pair of dictionaries $D_H$ and $D_L$ in the SR method by LD-WD depends on the high frequency patches from the training images set, which are randomly selected from each of the

images databases; both high resolution and low resolution images. For each patch of size 5, we subtract the mean pixel value. In all the experiments, we fix the dictionaries to the size $m \times 512$ where $m = 25$ for $D_H$ and $m = 100$ for $D_L$ and we selected 114 face images as training images from the Extended Yale-B database, which is described in the following section 4.7.

To estimate the effectiveness of the LD-WD method in terms of image reconstruction, we conducted experiments to measure and compare image quality of the super-resolved image by using PSNR measure. In these experiments, we also used SR by LD-Sp and the bi-cubic interpolation method to test whether improvement in image quality could be achieved by such existing methods.

It can be seen from Figure 4.8 that, at each level of degradation and wavelet decomposition, SR based wavelet dictionaries achieves superior improvements than interpolation technique. However, SR by LD-Sp in the spatial domain produces a slightly better image than LD-WD. Furthermore, it can also be seen that there is a small difference between all the wavelet filters. Therefore, the type of wavelet filter (filter length) has no effect on the reconstructed image; hence, we will use the shorter wavelet filter db2 and level 3-wavelet decomposition scheme in all the experiments. These results are consistent with the results of the super resolved image by using the iterative interpolation SR method in the spatial or wavelet domain.



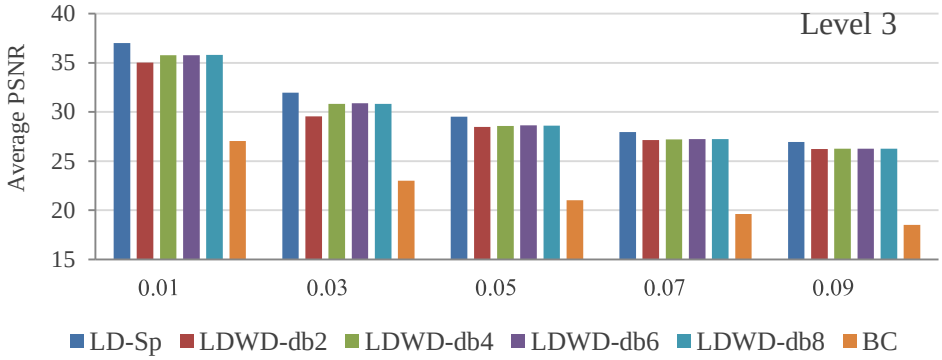



**Figure 4.8:** Comparison of PSNR values for super resolved images, between bi-cubic method & CS dictionary based SR method in the spatial/and wavelet domains.

The viability of SR method by LD-WD in terms of image quality was further tested by subjective visual comparison of the enhanced images by the learning dictionaries based super resolution methods in the spatial domain and in the wavelet domain. The comparison has been conducted on four LR test images shown in the second column of Figure 4.9. To simulate the intended uncontrolled scenarios, the LR images were generated from the good images by blurring them with Gaussian blurring degradation and down sampling. As can be seen form Figure 4.9, the quality of super-resolved images based on LD in the spatial domain is better than the one enhanced by LD in the wavelet domain. The images enhanced by dictionaries in the wavelet domain algorithm are with artifacts compared with the enhanced image in the spatial domain LD-Sp (see the cropped region of the first image shown). The analysis of the visual inspection of the images in Figure 4.9 is supported by the respective PSNR values.

In summary, the above experimental results shows the effectiveness of SR methods in the spatial domain (i.e. SR by LD-Sp and IISR-Sp), which produce better results, in terms of super resolved image quality, than enhancement techniques in the wavelet domain. Hence, SR methods in spatial domain will be used in all the experiments of this chapter.

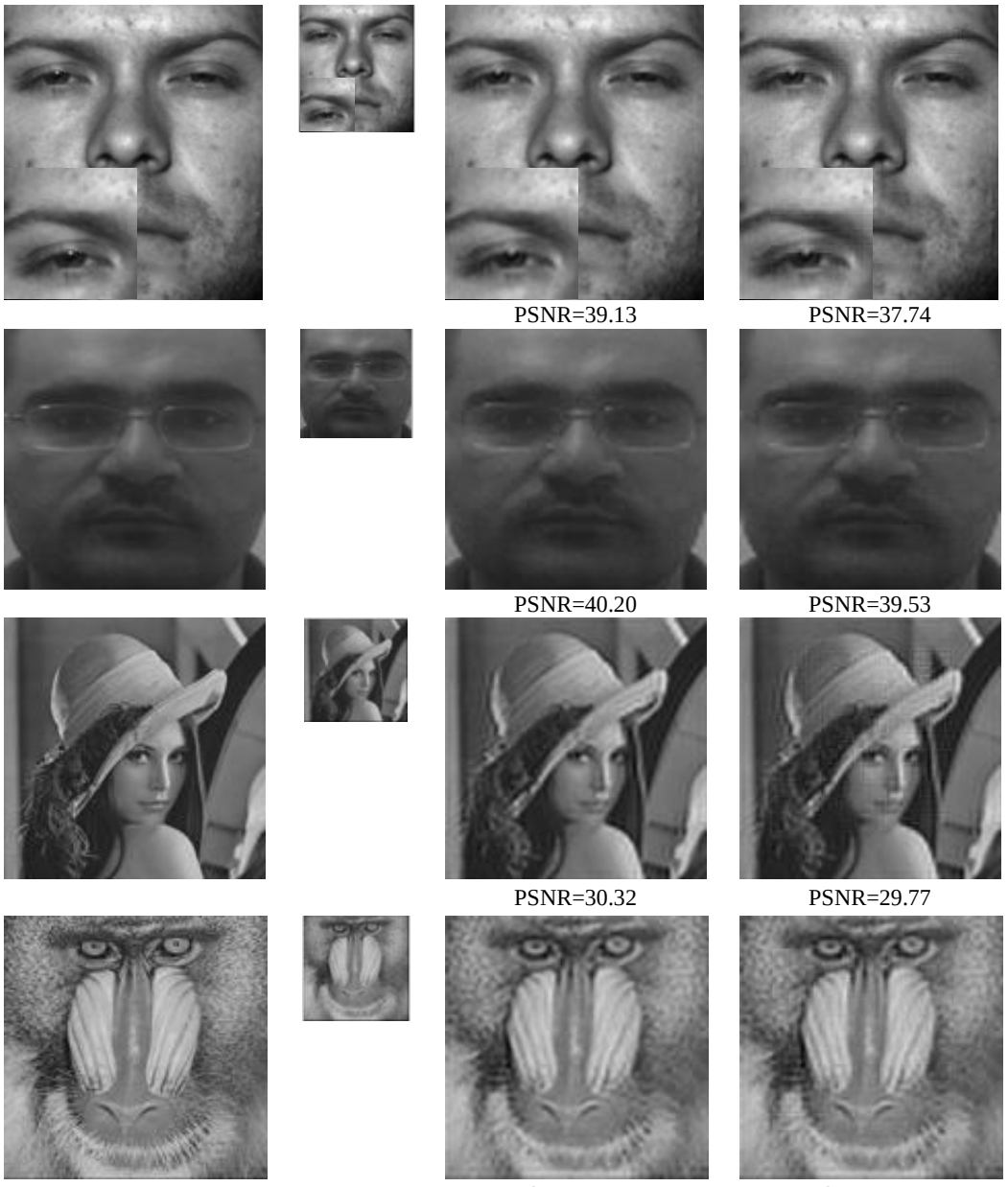



**Figure 4.9:** Comparisons of PSNR values between enhanced images, (a) First column: the original image. (b) Second column: the LR image generated by blurring and down sampling the original image. (c) Third column: super-resolved image by a factor of 2 using SR by LD-Sp method. (d) Fourth column: the super resolved image by a factor of 2 using SR by LD-WD method.

## 4.5 Novel Approaches to Construct Compressive Sensing Dictionaries

Compressive sensing based SR method, developed in (Yang, et al., 2010), have provided good evidences that the degraded LR image could be reconstructed by a sparse expansion in terms of suitable dictionaries built from large training images. This led us to investigate CS dictionaries in further detail with focus on the construction of new non-adaptive overcom-

plete dictionaries in such a way that could eliminate the need for a large training HR image set. In particular, we are interested in studying dictionaries that satisfy CS-related properties relevant to the recovery of sparse signal, such as those discussed in Chapter 3. In this section, we briefly describe our approaches for generating dictionaries that are independent of training images, but designed to implicitly be of full spark. These dictionaries will be used to recover SR images from a degraded LR image, with the aim of using them for face recognition at a distance – face recognition experiments will be discussed in the next chapter.

### 4.5.1 Iteratively Constructed Full Spark Dictionaries

Full-spark dictionary is a class of full row rank overcomplete $m \times n$ dictionary, where $m \ll n,$ so that each m-column's sub-matrix is a basis of $\mathbb{R}^m$. Here we describe an example of how to construct such matrices by starting with an invertible $m \times m$ matrix and iteratively appending a set of linearly independent $m$-column vectors in $\mathbb{R}^m$, while maintaining the full spark property after every addition. One way to maintain the full spark is to insist that every new column can only be generated by the full columns of the previously inserted sub-matrices.

Our generic full spark dictionary, referred to as LID, is of the form:

$$\boldsymbol{D} = \left[\boldsymbol{A_{p_1}}, \boldsymbol{A_{p_2}}, \dots, \boldsymbol{A_{p_k}}, \boldsymbol{C}(\boldsymbol{A_{p_{k+1}}})\right] \tag{4-5}$$

where for $i = 1, \dots, k$, the $p_i's,$ are distinct real numbers > 1, and

$$\boldsymbol{A_{p_i}} = \begin{pmatrix} \mathbf{1} & \frac{\mathbf{1}}{\boldsymbol{p_i}} & \frac{\mathbf{1}}{\boldsymbol{p_i}^2} \dots & \frac{\mathbf{1}}{\boldsymbol{p_i}^{m-1}} \\ \frac{\mathbf{1}}{\boldsymbol{p_i}} & \mathbf{1} & \frac{\mathbf{1}}{\boldsymbol{p_i}} \dots & \frac{\mathbf{1}}{\boldsymbol{p_i}^{m-2}} \\ & & \vdots & \\ \frac{\mathbf{1}}{\boldsymbol{p_i}^{m-1}} & \frac{\mathbf{1}}{\boldsymbol{p_i}^{m-2}} & \frac{\mathbf{1}}{\boldsymbol{p_i}^{m-3}} \dots & \mathbf{1} \end{pmatrix} \tag{4-6}$$

Note that $k \cong n/m$ and the last sub-matrix of $D$ is an $m \times r$ matrix with $r < m$. Then, the $m \times n$ LID dictionary is obtained from the resulting matrix after normalising its columns using the $l_2$-norm.

For experimental purposes, the $\text{LID}_1$ high-dictionary $D_H$ is generated using integer numbers $p_i > 1$ and the other $D_H$ dictionary, which is denoted by $\text{LID}_2$, is generated using real numbers between $1 < p_i \leq 2$. For simplicity, the low-dictionary $D_L$ was created from a Standard Gaussian Random Matrix (SGRM).

### 4.5.2 Constructed RIP Orthonormal Dictionaries

This is a class of overcomplete dictionaries in $\mathbb{R}^{m\times n}$ where $m \ll n$, which satisfy RIP by design where the ideal well-conditioned dictionary can be constructed by concatenated pairs of orthonormal matrices $U$ & $V$ of size $m \times m$ and $n \times n$ respectively, and the non-zeros diagonal entries of these matrices is a $2 \times 2$ rotation matrix (RM), otherwise are zeros. For randomly selected sets of thetas, $\theta = \{\theta_1, \theta_2, .., \theta_r\}$ in the range [0, $2\pi$], the rotation matrix can be defined as:

$$\boldsymbol{RM} = \begin{bmatrix} \boldsymbol{\cos\theta} & \boldsymbol{\sin\theta} \\ \boldsymbol{-\sin\theta} & \boldsymbol{\cos\theta} \end{bmatrix} \tag{4-7}$$

While, the orthonormal matrices $U$ and $V$ are represented as:

$$\boldsymbol{U} = \begin{bmatrix} \boldsymbol{RM_1} & \boldsymbol{0} & \cdots & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{RM_2} & \cdots & \boldsymbol{0} \\ \vdots & \vdots & \cdots & \vdots \\ \boldsymbol{0} & \boldsymbol{0} & \cdots & \boldsymbol{RM_m} \end{bmatrix} \quad \text{and} \quad \boldsymbol{V} = \begin{bmatrix} \boldsymbol{RM_1} & \boldsymbol{0} & \cdots & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{RM_2} & \cdots & \boldsymbol{0} \\ \vdots & \vdots & \cdots & \vdots \\ \boldsymbol{0} & \boldsymbol{0} & \cdots & \boldsymbol{RM_n} \end{bmatrix}$$

Therefore, the new proposed dictionary denoted by ROM can be described as the form:

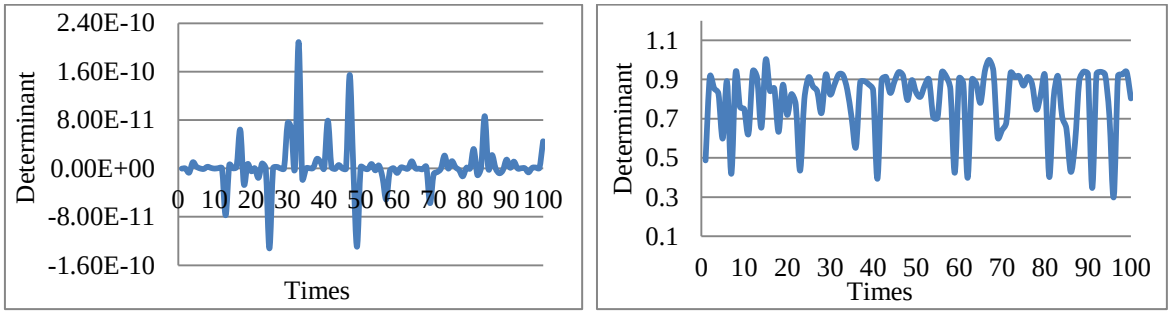
$$\boldsymbol{ROM} = \boldsymbol{U\Sigma V^T} \tag{4-8}$$

Where Σ is an $m \times n$ well-conditioned matrix with non-zero diagonal values and zero elsewhere, the diagonal entries are randomly selected in decreasing order, such that the ratio between the maximum and minimum value is less than 2.5. Note that this construction is based on the use of singular value decomposition, and the condition on the ratio of maximum and minimum coefficients in the diagonal matrix is a requirement to ensure the satisfaction of the RIP condition (see chapter 3).

## 4.6 Comparisons of RIP parameters for Different Dictionaries

In this section, the comparisons of "strength" of the sufficient CS conditions for the proposed and existing dictionaries such as various random dictionaries and learnt dictionaries will be illustrated. The first set of comparisons we conducted were for the "strength" of RIP for the LD-Sp, $LID_1$ and ROM dictionaries, based on statistical testing for NSP. This is due to the fact that an exhaustive search is not feasible since there are $\binom{512}{25} = \frac{512!}{25!(512-25)!}$ sub-matrices of the dictionary. Hence, we opted to evaluate the determinants as indicators of linear independence, for a hundred randomly selected sub-matrices. The results are displayed in Figure 4.10

below. Although in theory, the LD-Sp dictionary may satisfy NSP, practically and computationally this is not the case because the determinants of most of these sub-matrices are close to zero. In contrast, the results of all the determinant values of the example $LID_1$ dictionary are comfortably away from zero, as well as the determinant value for non-zero m-columns in the sparse ROM dictionary being > 0.034, which indicates the proposed dictionaries $LID_1$ and ROM statistically satisfy NSP.




(a) (b)

**Figure 4.10:** Determinant of a hundred sub-matrices from: (***a***) Learning Dictionary (***b***) Proposed $LID_1$ Dictionary.

To follow up on the above observation that the results from the LD-Sp dictionary may be somewhat ill-conditioned, we conducted a similar statistical experiment as above but this time calculating the condition numbers of the randomly selected sub-matrices for 5-dictionaries ROM, GRM, TCRM, $LID_1$, and LD-Sp. These condition numbers are expected to be bounded by RIC of order $2k$, with $k$=12. Table 4.1 shows the per dictionary mean and standard deviation of the condition numbers for a hundred randomly selected sub-matrices of different sizes, ranging from $25 \times 25$ to the full size of $25 \times 512$.

| submatrices | Different Dictionaries | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | $LID_1$ | | LD-Sp | | ROM | | GRM | | TCRM | |
| | mean | std | mean | std | mean | std | mean | std | mean | std |
| 25x25 | 2.09 | 1.73 | 3.34E+16 | 1.79E+17 | 1.24 | 0.33 | 277.61 | 1.11E+03 | 85.19 | 155.68 |
| 25x50 | 1.80 | 0.93 | 1.42E+15 | 1.79E+14 | 1.19 | 0.14 | 4.96 | 0.61 | 4.34 | 1.38 |
| 25x75 | 1.71 | 0.68 | 1.20E+15 | 9.16E+13 | 1.19 | 0.14 | 3.25 | 0.24 | 2.85 | 0.32 |
| 25x100 | 1.65 | 0.55 | 1.14E+15 | 5.69E+13 | 1.19 | 0.14 | 2.59 | 0.14 | 2.60 | 0.25 |
| Full matrix | 1.84 | | 1.00E+15 | | 1.48 | | 1.43 | | 1.45 | |

**Table 4.1:** The mean and standard deviation for condition numbers, for a hundred random sub-matrices of different sizes.

This is an added indicator of the successful recovery of sparse vectors. These results again demonstrate that the $LID_1$ and ROM are well-conditioned matrices in comparison to all other dictionaries for the various sub-matrices. Although there is no efficient algorithm to test the sufficient RIP property (Li, et al., 2012) of the random matrices, the condition number for the full matrix indicated that the GRM and TCRM matrices have small and similar condition numbers as well as are slightly better than the constructed dictionaries.

Moreover, the condition number of the LD-Sp is extremely large for all cases, which makes these dictionaries very ill conditioned. In fact, $LID_1$ and ROM are well conditioned dictionaries as a full matrix where the condition number is small and equal to 1.84 which is less than 2.5 while the condition number for the full LD-Sp dictionary is large, nearly 1.00E+15, which makes this dictionary highly ill-conditioned.

Furthermore, we conducted similar statistical tests, but this time of coherence numbers. Figure 4.11 illustrates that for the LD-Sp dictionary, all the $\mu$-coherence of the sub-matrices are very small and slightly outside the lower boundary $\mu \geq \frac{1}{\sqrt{m}} = 0.2$, which implies that this dictionary goes far to satisfy the RIP. In comparison, the coherence values for almost all sub-matrices from $LID_1$ and random matrices are comfortably inside the boundary $[\frac{1}{\sqrt{m}}, 1]$.

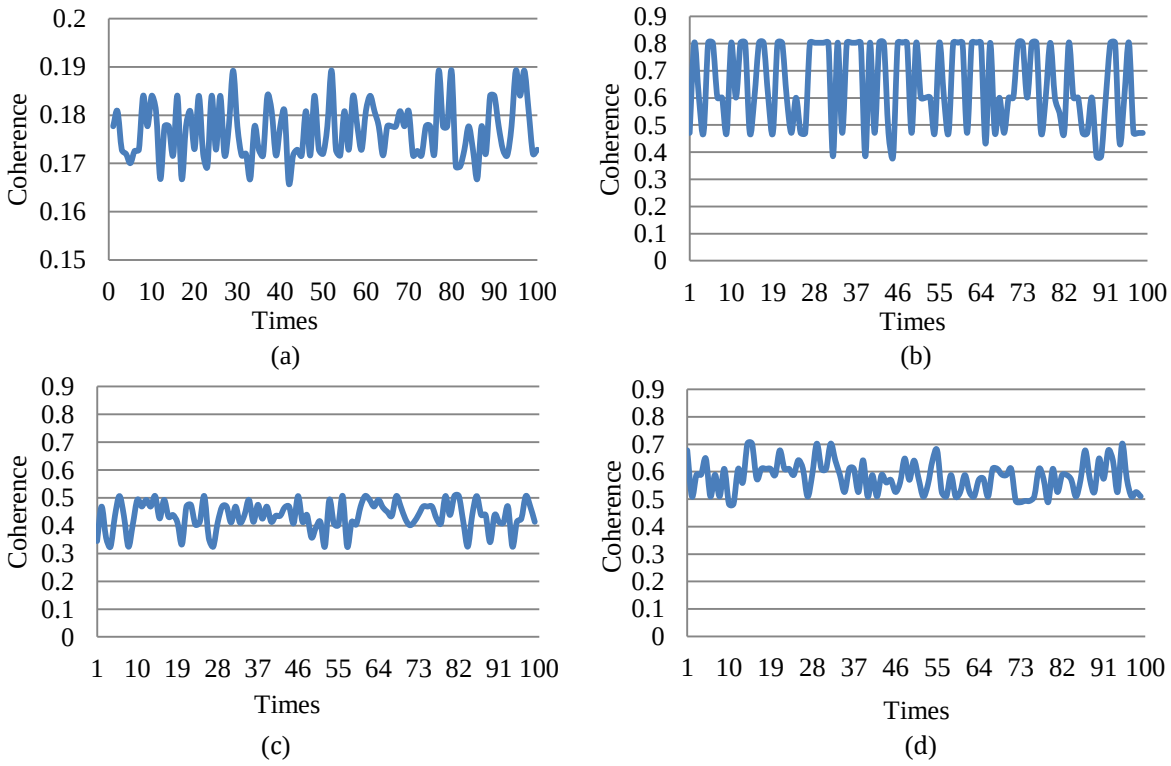



**Figure 4.11:** Mutual Coherence for sub-matrices from: (a) LD-Sp dictionary, (**b**) $LID_1$ dictionary, (**c**) TCRM dictionary and (**d**) GRM dictionary.

Finally, the coherence values for the full dictionaries as well as their spark are displayed in Table 4.2. It is well known that the highest sparsity recovered signal for any dictionary equal (1+row rank)/2.

| CS Properties | Different Dictionaries | | | | |
|---|---|---|---|---|---|
| | $LID_1$ | LD-Sp | ROM | GRM | TCRM |
| Row Rank | 25 | 24 | 25 | 25 | 25 |
| Coherence-all matrix | 0.9958 | < 0.2 | 0.2553 | 0.7438 | 0.7318 |

**Table 4.2:** Spark and Coherence Properties for Different Overcomplete Dictionaries.

The results demonstrate that the constructed dictionaries $LID_1$ and ROM have a large value of spark with a row rank equals to 25 as the random matrices, which implies that there are no $m$-columns that are linearly dependent on the dictionaries, whilst the LD-Sp dictionary has a high probability of finding linearly dependent columns where a row rank equals 24.

## 4.7 Database Description

This section gives a brief description of two publicly available face biometric databases that are used in our experiments to evaluate the suitability of the different dictionary methods based SR, including the proposed dictionaries as well as state-of-the art methods, to super resolve face images from a single or set of LR images with two models of blur degradation.

The first database is ***UBHSD video database*** (Al-Obaydy & Sellahewa, 2011). This database contains 20 subjects from both genders recorded by different quality video cameras: a high-definition (HD) camera and a standard-definition (SD) camera. The videos of each subject were recorded in two sessions with a gap of at least two days between the recording sessions. Each recording session includes videos captured in controlled indoor conditions (videos captured in the same room under semi-controlled lighting with a uniform background), and in uncontrolled outdoor environment. In each recording session and at each location (indoor location and outdoor location), three image frames of each subject were obtained at four different ranges/distances from the camera, where range 1 ($R_1$) being the nearest to the camera, while range 4 (R4) is the farthest from the camera. Thus, twelve frames are selected to capture a subject at four distance ranges. The UBHSD database consists of blurred face images, faces with eyes closed and slightly varying poses, and includes head movements and facial expressions whilst the subjects walk in a natural way. Each subject has 96 face images.

Hence, the total number of face images in the database is 1920. All face images are converted to gray scale and resized $128 \times 128$ pixels. Figure 4.12 shows examples of face images from the UBHSD database.

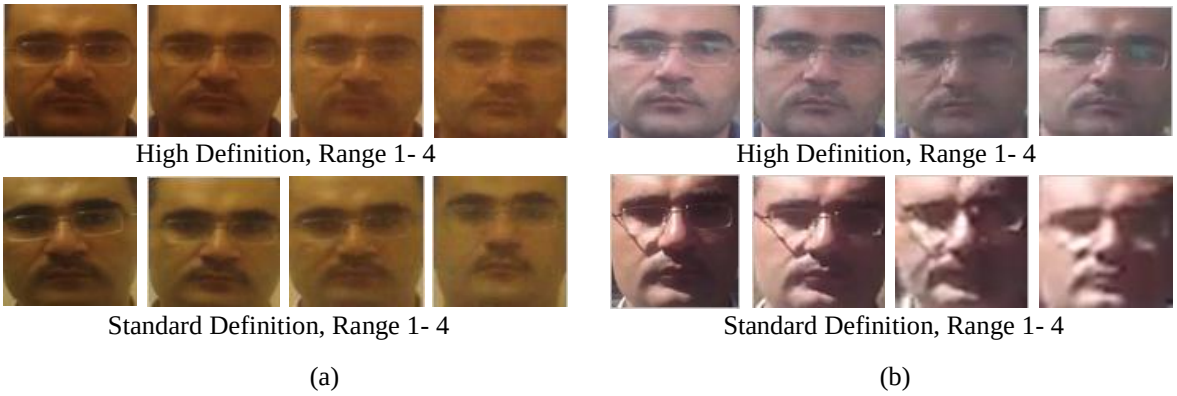



**Figure 4.12:** Examples of cropped and rescaled face images from HD and SD videos captured in: (a) Indoor Condition (b) Outdoor Condition.

The second database is the widely used ***Extended Yale B database*** (Lee, et al., 2005), (Georghiades, et al., 2001). This database consists of $2414$ face images of $38$ individuals each having 64 images, in frontal pose. The cropped and normalized face images of size $192 \times 168$ were captured under various laboratory-controlled lighting conditions. The images in the database are divided into five illumination subsets according to the angle $\theta$ of the light-source with respect to the optical axis of the camera as shown in Table 4.3.

| Subsets | Angles | No. Of Images |
|---|---|---|
| 1 | $\theta < 12$ | 263 |
| 2 | $20 < \theta < 25$ | 456 |
| 3 | $35 < \theta < 50$ | 455 |
| 4 | $60 < \theta < 77$ | 526 |
| 5 | $85 < \theta < 130$ | 455 |

**Table 4.3:** Different illumination sets in the extended Yale B database

We selected three images per subject from subset one, which contain images of good quality, as HR images to construct the learning dictionary for super resolved images. The total number of HR images in the training set for the dictionary is 114 and includes images of both genders. Figure 4.13 shows some examples of training face images taken from the Extended Yale B database.

**Figure 4.13:** Example of three out of 114 training face images.

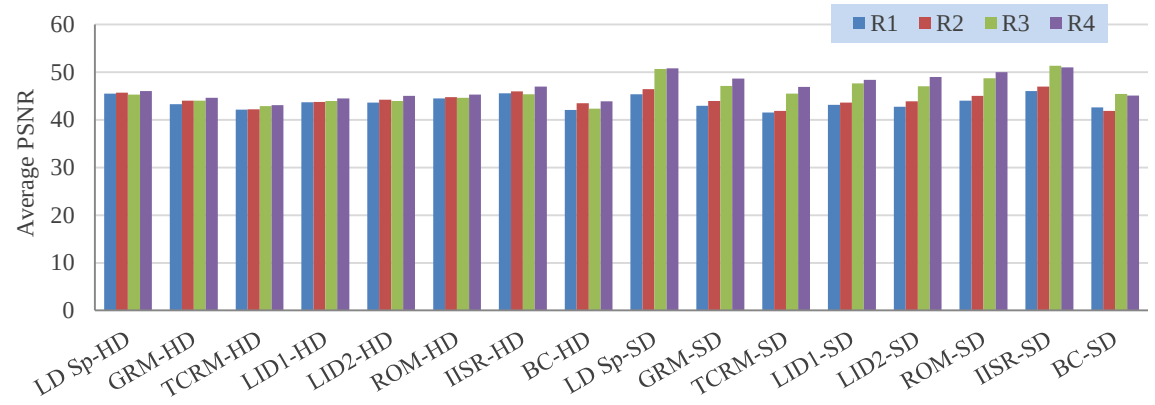
## 4.8 Image Quality Evaluation and Discussion

The aim of this section is to test the performance of various CS dictionaries in super resolving image. In general, super resolution approaches are expected to produce better quality images than the well-known interpolation method. Therefore, we compared the effectiveness of the various SR dictionary methods, the IISR method and the bi-cubic method by measuring the quality of their super resolve images using variety of quality measures. Two groups of experiments have been conducted and are reported in this section.

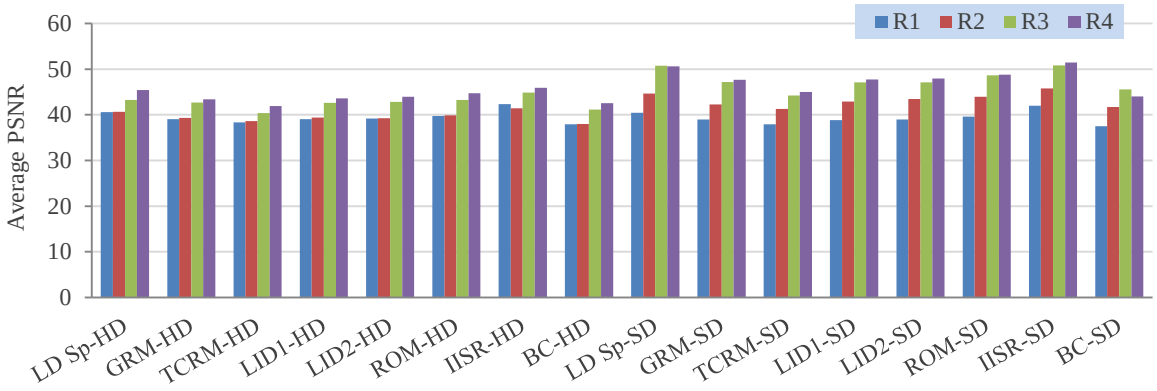
### 4.8.1 Experiment 1: Results of UBHSD video Database

We conducted experiments to measure and compare image quality in the super-resolved images with respect to the original good quality images in the data set, using PSNR measure quality. The original face images of size $128 \times 128$ were down-sampled to size $64 \times 64$ to create LR test images. Note that, the LR images in this database not generated from any model of blurring degradation. The test images were super-resolved using various SR techniques. The original images of the UBHSD database were used as the ground truth to measure the quality of super-resolved images.

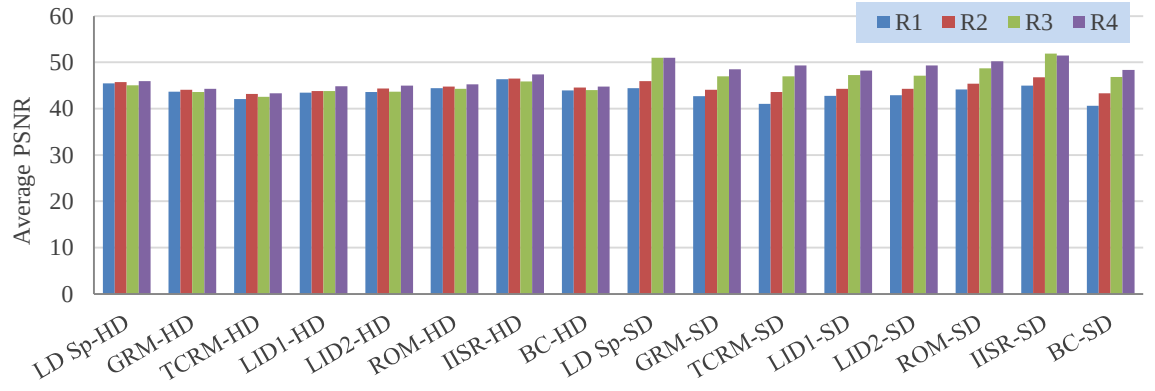
It can be seen from Figure 4.14 that, at each distance range, for each session and in different conditions (indoor and outdoor), the various CS dictionaries based SR as well as the non-CS based iterative interpolation SR method produce better quality improvement than the standard bi-cubic interpolation method. However, there is only a small difference in terms of image quality between the non-adaptive (i.e., independent of the face images) CS dictionary methods such as LID1, LID2, GRM, TCRM, and ROM that. On the other hand, the image learnt dictionaries that depend on training images (LD-Sp) as well as the iterative SR method are able to achieve slightly better results than the non-adaptive CS dictionaries, but this improvement is not significant. Hence, generally, the proposed non-adaptive CS dictionaries as well as the random matrices are able to improve the quality/resolution of the LR images as well as the learning dictionaries based SR method.



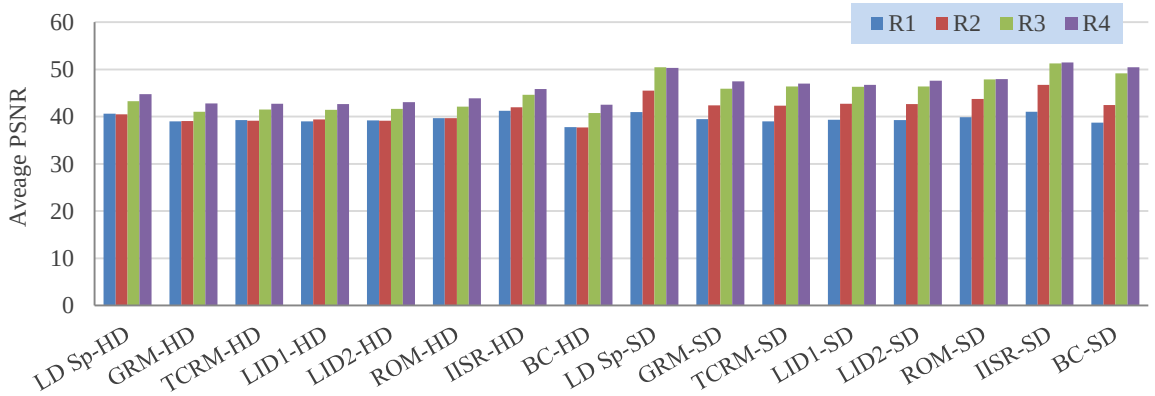

HD and SD video data from indoor, session 1




HD and SD video data from outdoor, session1




HD and SD video data from indoor, session 2




HD and SD video data from outdoor, session 2

**Figure 4.14:** A comparison of PSNR values for super resolved images in the different distance ranges.

### 4.8.2 Experiment 2: Results of Extended Yale B database

The viability of various dictionary based SR techniques to reconstruct super resolved images from any degraded LR face images are tested in this section. Two sets of experiments were conducted on the frontal face images from the Extended Yale B database, where the degraded LR images were generated by blurring the good quality images in set 1 and set 2 using one of two models of degradation followed by down sampling by a factor two. In the first set of the experiments, Gaussian filter of size 7 will be used to model blur degradation. However, the Gaussian function in general has a low level of blurring effect on an image. Therefore, in the second set of the experiments, atmospheric turbulence degradation function with different strengths, which has a complex degradation effect on an image, reflect the degradation visible in uncontrolled environment conditions such as wind speed, and captured at a distance, will be used to model blur degradation.

**Low Resolution Image with Gaussian Blurring Degradation**

The first stage of the experiments is to test the performance of SR method by LD-Sp that relying on training images, IISR method and bi-cubic technique on all the face images in the Extended Yale B database. Figure 4.15 illustrates average PSNR values where $LR_1$ and $LR_2$ are the low-resolution images obtained by taking even and odd indexed pixels respectively from blurred original images with Gaussian filter. From Figure 4.15 we can observe that the SR method by LD-Sp produce superior improvement in terms of PSNR when compared to the standard bi-cubic (BC) interpolation method. However, the performance of iterative interpolation SR method slightly outperforms the SR by LD-Sp for each subset of the Extended Yale B database. Furthermore, the performance of the SR methods in subset 5 is slightly higher than other sets in the database due to the effect of the blur on the poor illumination quality images in subset 5, which is either very little or there is no effect, hence the illumination seems to remain an added challenge. This will be part of future investigations and will be linked to face recognition tasks. Therefore, we will use SR methods to recover super resolve images from LR images with well lit in set 1 and set 2.

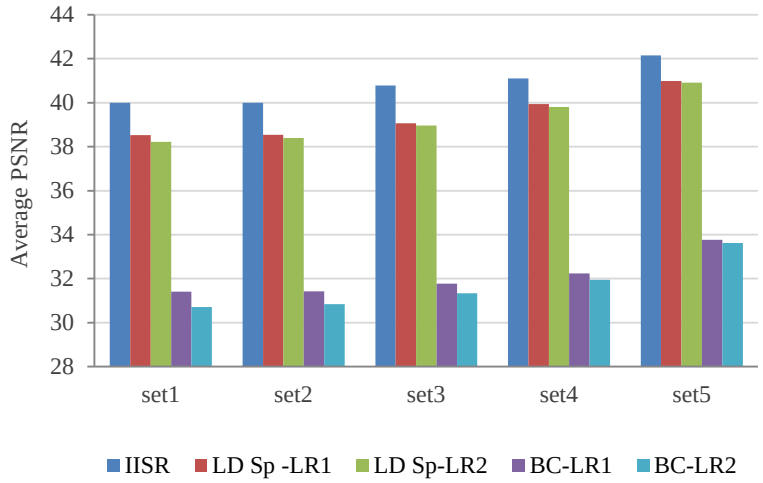



**Figure 4.15:** PSNR values for super-resolved images from LR faces in the Extended Yale B database.

The second stage of the experiments used contrast and correlation quality measures (described in Section 1.4) as well as PSNR to evaluate the performance of SR by various types of dictionary methods and compare the results with the well-known interpolation method and IISR method. Figure 4.16 shows the results on the reasonably lit face images in sets 1 and 2 of the database. The results demonstrate that for each quality measure, the bi-cubic method produced a lower quality image than the dictionary-based methods. Moreover, there is no significant difference between LD dictionary in the spatial domain and the various dictionaries constructed without using training image information. Furthermore, the performance of SR by different dictionaries produced the same than or slightly better quality image than the iterative interpolation SR method. Hence, from all the values of the different quality measures led us to the conclusion that the SR by non-adaptive CS dictionary methods that do not depend on training images and the SR by LD-Sp are able to achieve superior improvement in terms of image quality than the standard interpolation method.

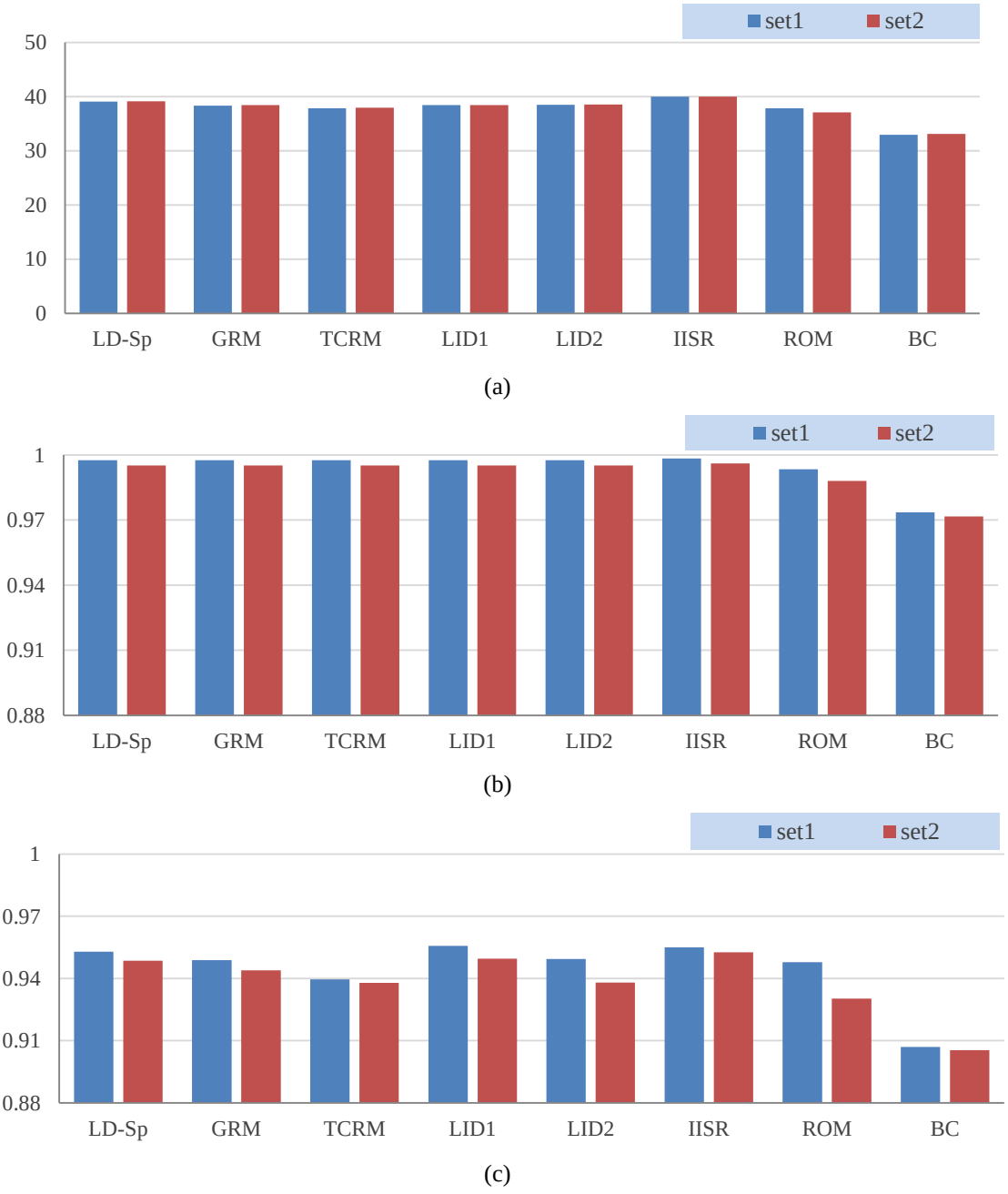



**Figure 4.16:** Average values of (a) PSNR measure (b) Contrast measure, and (c) Correlation measure, for the super resolved images that reconstructed using different dictionary methods, IISR method, and interpolation method.

Note that the above quality measures require a reference image to calculate the quality of a given image. However, in operational situations, the original HR image that was used as a reference to calculate the above quality measures is not available or very hard to obtain (there would be no need to super-resolve a LR image if the original HR image was available). For this reason, we used an adaptive (i.e., no-reference) quality measure, known as "*histogram intersection"* (HI), which uses the histogram of a wavelet subband of an image to measure its

quality. In the third stage of the experiments, the histogram intersection will be used to test the performance of different overcomplete dictionary methods and compare the results with the non-CS based iterative interpolation method and the interpolation technique. First, we explain the histogram intersection based image quality measure.

Histogram intersection (AL-Jawad, 2009) is a quality measure between the actual histogram (AHistogram) and theoretical Laplacian distribution histogram (THistogram) of the wavelet subband coefficients define as the sum of the corresponding minimum histogram values:

$$\mathbf{HI} = \sum_{i=1}^{N} \mathbf{min}(\mathbf{AHistogram}(i), \mathbf{THistogram}(i)) \tag{4-9}$$

Where THistogram is calculated based on the standard deviation (STD) of the actual histogram as shown below:

$$\mathbf{THistogram}(x) = \frac{1}{2\sigma} e^{(-\frac{|x-\mu|}{\sigma})} \tag{4-10}$$

$\sigma$ =STD (Actual Histogram) and $\mu$ =Mean (Actual Histogram), where $\mu \cong 0$. AL-Jawad in (AL-Jawad, 2009) used HI as a quality measure for images that are degraded by varying lighting conditions and demonstrated that the HI values in the LH wavelet subband give a good indication whether the image is nearer to the ideal well-lit condition or not. For the well-lit images (i.e. images from set 1 or set 2), the AHistogram is nearer to the THistogram, while for darker images the AHistogram is further away from THistogram. In our work, we use HI to measure image quality for the super resolved images. The wavelet transform has been applied to level one to extract HI values. The theoretical Laplacian distribution (4-10) has been calculated based on the STD of non-LL subband (LH) and will be compared with the actual one.

We can observe from the Figure 4.17 that:

- the intersection value for the enhanced images by interpolation method and iterative SR method reached around to 6.72 and 10.64 respectively, while the intersection value for the super resolved images for example by the proposed ROM dictionary reached to 12.85. This demonstrates that the performance of various types of dictionaries outperforms the standard BC interpolation technique and IISR method.
- the histogram intersection values reveal that the proposed dictionaries such as $LID_1$, ROM, and $LID_2$ and the random dictionaries GRM and TCRM that are independent of image data are able to improve image quality and produce the same, if not slightly better, results than LD-Sp that depends on images data.

- The results illustrate there is a small difference between the different dictionaries that constructed without using image information.

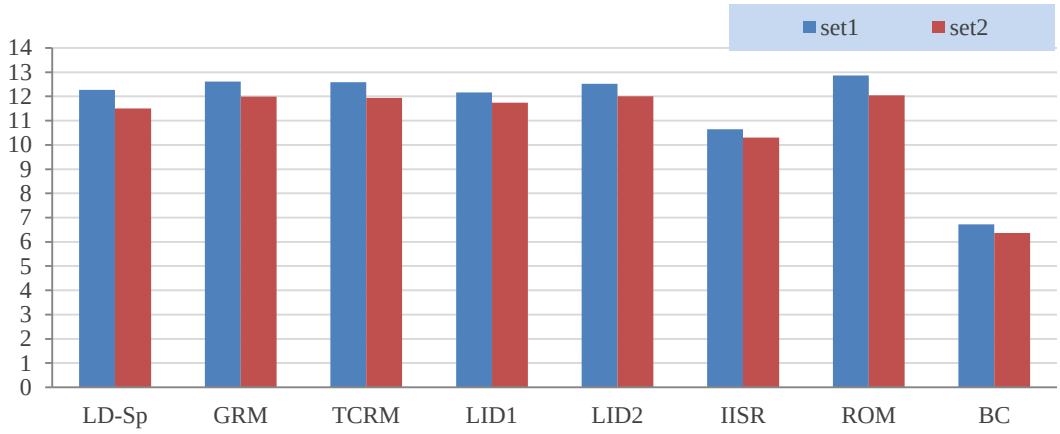



**Figure 4.17:** Average values of the histogram intersection quality measure for the super resolved images that are recovered using different CS dictionary based methods, non-CS based iterative SR method and bi-cubic method.

Overall, we can conclude from the above experiments that the improvement in resolution comes from the proposed CS dictionary methods based SR rather than bi-cubic method and IISR method as well as produce slightly better image quality than SR method by learning dictionary. Moreover, the choice of dictionary had no significant impact on the quality of the super-resolved images when super-resolving images with low degradation (i.e., images with Gaussian degradation).

These results led us to the following question; could the non-adaptive CS dictionary methods as well as the learning dictionaries reconstruct image resolution from degraded LR image with sever blur degradation?. The next stage of experiments in our work aims to answer to this question. The non-CS iterative interpolation SR method and bi-cubic interpolation will be used as a benchmark for comparisons.

### Low Resolution Image with Turbulence Degradation

Dictionary based SR methods to super-resolve image with good quality from degraded LR image with different level of degradation was tested. The LR test images were generated from the images in set 1 and 2 of the Extended Yale B database by blurring the original image using Turbulence function followed by down sampling. The reconstruction reliability is quantified using PSNR, Contrast, Correlation, and histogram intersection. Figure 4.18 pre-

sents the average quality obtained by each of the SR methods for a reasonably wide set of blurring degradations, ranging from minor degradation corresponding to $k_1$=0.01 to severe degradation corresponding to $k_1$=0.09. We can see from Figure 4.18 that:

- at each level of degradation, there is no significant difference in the quality of images recovered by various types of dictionary methods;
- the various dictionary methods are able to improve image quality and produce better results than the bi-cubic interpolation technique. However, histogram intersection quality values presented slightly different indication for severe degradation (i.e. $k_1$>0.07), where bi-cubic technique produced slightly better quality image than the SR methods, but this improvement is not significant.
- Based on the histogram intersection values, the various types of dictionary methods produce superior improvement than the non-CS based IISR method for all levels of degradation, including no blurring.
- unsurprisingly, and regardless of the method used in the SR procedure, all measured values of super resolved image quality decrease as the level of blurring increases, and
- the PSNR, Contrast, and Correlation quality values of the images enhanced by bi-cubic technique decrease with increasing level of degradation. However, the histogram intersection values are slightly varied where the quality of the images enhanced by bi-cubic increases with increasing level of degradation in the LR images.

In general, the experimental results illustrate that the CS dictionary based SR method are able to reconstruct image resolution and image quality from any degraded LR images and yield significant improvement than the non-CS based IISR method and the well-known interpolation technique.

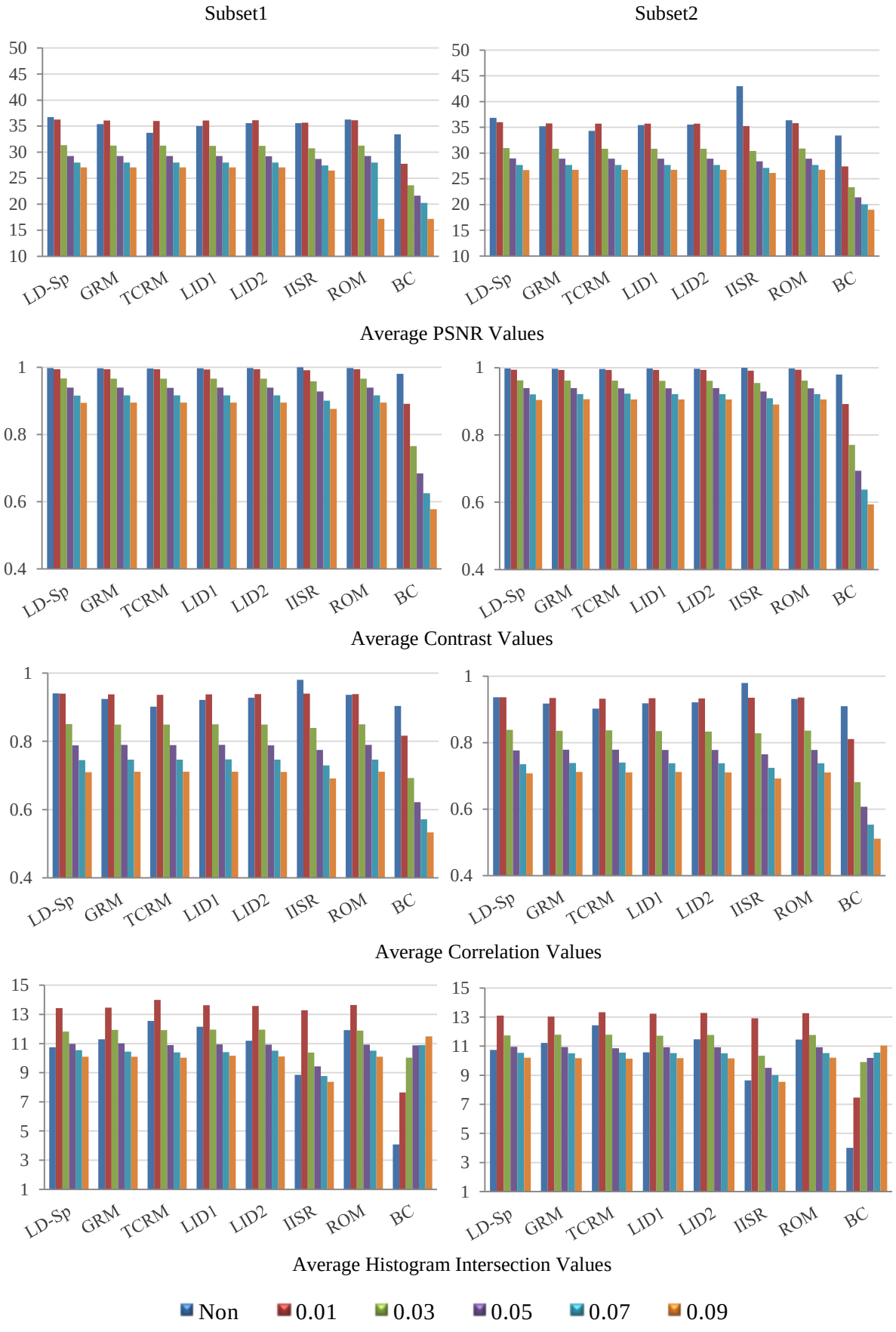



**Figure 4.18:** Comparisons of different measured values for super resolved images in the Extended Yale B database.

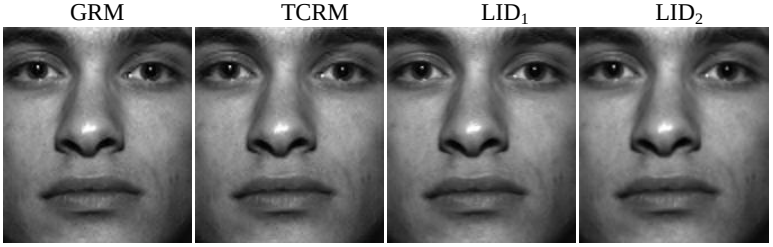
Finally, in order to test the level of success of the CS-based SR schemes, we conducted a subjective comparison between enhanced images by the various dictionary methods, the non-dictionary iterative SR method, and bi-cubic interpolation method. Moreover, wiener restoration technique (see chapter 2) to remove possible blur degradation in the enhanced image by bi-cubic interpolation method was used for the comparisons. Figure 4.19 shows an example of an original HR image, its degraded and down sampled versions, and the super resolved images using the various schemes. The degraded LR images were obtained from the HR image by applying the degradation function in equation (4-3) for different $k_1$values ranging from low to severe, followed by down sampling. As can be seen in Figure 4.19 below, the improvement can be noticed in the SR images by CS dictionary methods, including LD-Sp and in the IISR method over the LR images and the standard interpolation method as well as over the restored interpolated image. Furthermore, at every level of degradation, the proposed dictionaries and the random matrices can produce good quality images similar to, or slightly better than, those produced by dictionaries that rely on training image data.

Overall, the difference between the reconstructed HR images using various dictionaries is not discernible by the human eye and the images are quite similar to each other. This leads to the conclusion that the training image set is not necessary to create the dictionaries for image super resolution.

Original HR image

GRM TCRM $LID_1$ $LID_2$

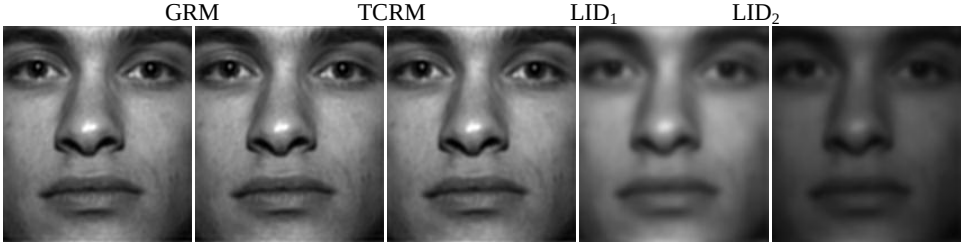
LR, no blur-ring



LD-Sp ROM IISR BC

GRM TCRM $LID_1$ $LID_2$

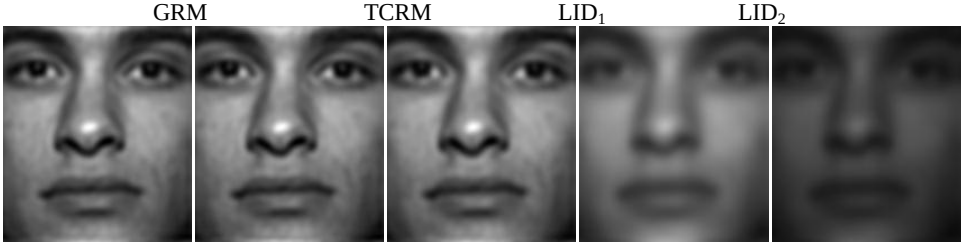
LR, $k_1$=0.01



LD-Sp ROM IISR BC Restoration BC-image

GRM TCRM $LID_1$ $LID_2$

LR, $k_1$=0.03



LD-Sp ROM IISR BC Restoration BC-image

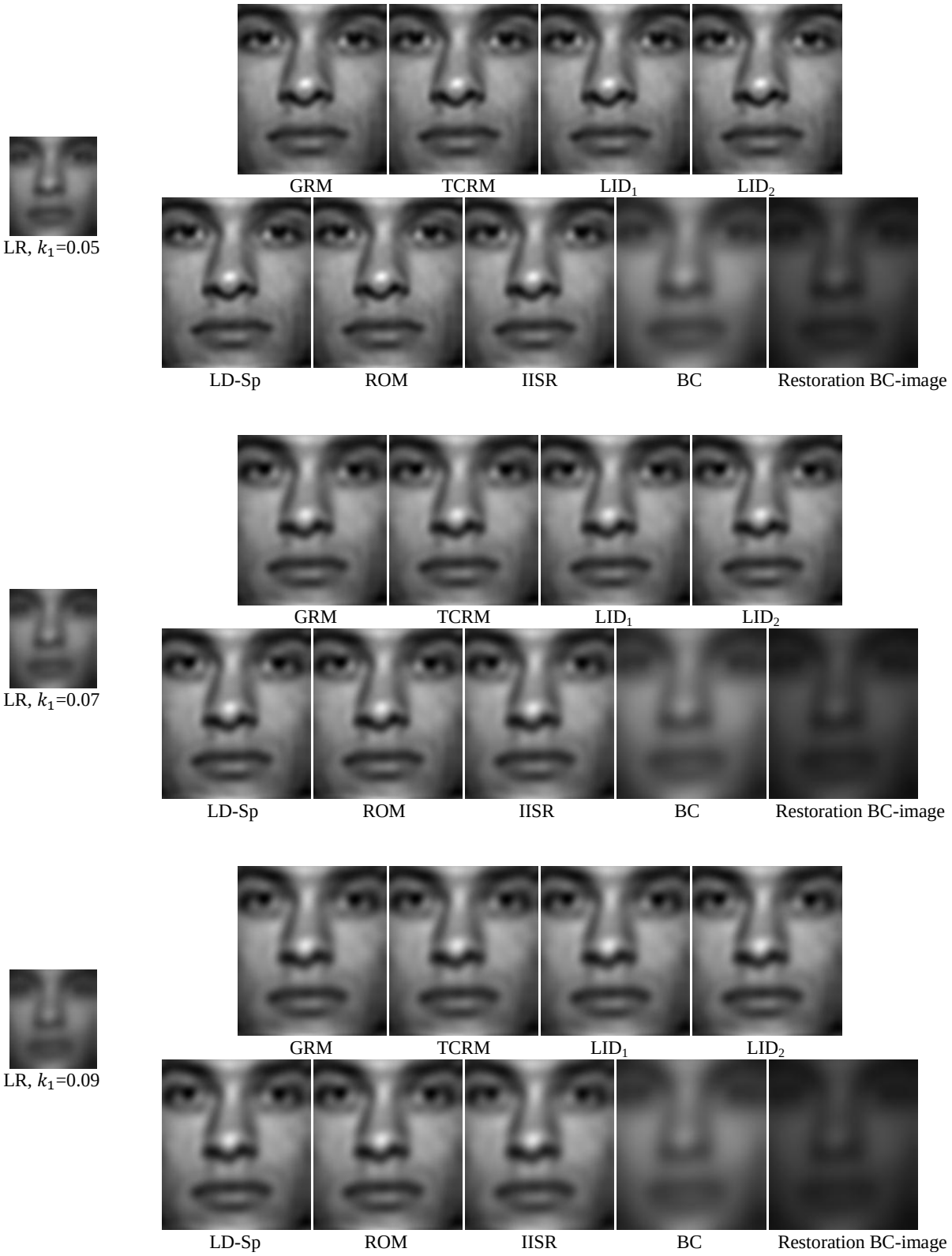



**Figure 4.19:** Comparison between SR approaches by different dictionaries, non-dictionary iterative method and bi-cubic interpolation method.

## 4.9 Summary and Conclusion

We investigated and tested the performance of CS dictionaries such as random dictionaries, dictionaries created using a set of HR training face images, and the proposed overcomplete dictionaries constructed based on CS properties and are independent of training image data, to super resolve degraded low-resolution images with different models of degradation. In addition, the viability of the SR techniques in the wavelet domain to super-resolve images was studied.

The results effectively support the use of SR based techniques that employ CS dictionaries, including the non-adaptive dictionaries that satisfy CS theory over the well-known interpolation technique and non-CS based IISR method to super-resolve images. More importantly, there is no need to use image sets to construct dictionaries, because the non-adaptive dictionaries perform equally well, if not better, than the learning dictionaries that depends on image information. Moreover, a visual inspection of the super resolved images using the various SR procedures reveals a significant improvement in the quality of the super resolved images compared to using the BC method.

We investigated the RIP property for various underdetermined matrices, and in order to find possible explanations for those matrices to satisfy the property of the CS, we conducted a number of tests of numerical matrix parameters relevant to the RIP condition. We noted that the learning image-based dictionaries are highly ill conditioned matrices and are far from satisfying the RIP related conditions discussed in the literature. Perhaps the use of training images (image patches) with the same statistical nature of the input face image (image patches of the LR image) compensate for the lack of RIP property. In the next chapter, we shall evaluate the performance of our dictionary methods for face recognition in terms of identification accuracy rates.

# Chapter 5

## FACE RECOGNITION FROM DEGRADED LR IMAGES

Face recognition under uncontrolled conditions, e.g. in distance surveillance scenarios and post-rioting forensics, whereby captured face images are severely degraded or blurred and of low-resolution, forms a tough challenge due to a range of factors. These factors include, but not limited to, difficulties in determining a model for image degradation that encompasses a range of realistic capturing conditions as well different qualities of cameras. The investigations conducted in Chapter 4 have demonstrated the importance and suitability of the recently developed compressive sensing based super resolution methods for recovering images with good resolution and quality. This chapter extends the work in last chapter by focusing on the use of compressive sensing (CS) theory to develop scalable face recognition schemes that do not require training, by using, various overcomplete dictionaries that were used in the previous chapter to construct a super resolve face and feature image from any degraded face image. It shall be demonstrated that the non-adaptive image-independent, implicitly designed, dictionaries as well as a variety of random dictionaries for the recovery of unique $k$-sparse signal, can yield significant face recognition accuracy rates that are as good as if not better than those achieved by the recently proposed image-based learnt dictionaries. This will remove the need for training images in developing face recognition schemes. However, we shall also reveal a problem that is associated with current approaches to the design of non-adaptive dictionaries in that they may not work well in super-resolving wavelet feature vectors in comparison to the way they succeed in super-resolving the corresponding spatial domain signals. In contrast, learnt dictionaries can be built in the wavelet domain and perform well. For comparison, we also test the performance of a non-CS based iterative SR method, bi-cubic interpolation method and matching in low-resolution (i.e. no super resolution is applied to LR degraded face images are matched against HR gallery).

## 5.1 Existing Works in the Literature

Face recognition in uncontrolled conditions arises primarily in fighting crime through surveillance using CCTV cameras. In contrast to recognition in controlled conditions, which has witnessed significant improvement over the last three decades, very little progress has been made in attaining acceptable recognition rates due to the degraded nature of the CCTV captured images (Grgic, et al., 2011). CCTV cameras are at a distance from the imaged scenes and therefore capture low-resolution, blurred low-quality face images, with variable lighting and pose conditions.

*In any face* recognition systems, one need to extract a set of feature vectors (one or more for each of the enrolled subjects) called the gallery, representing the digital templates that are obtained during the enrolment stage. For recognition under controlled conditions, these templates are extracted from good quality images of reasonable resolution, and when a claimant presents him/herself, a reasonably good quality image using similar devices to those used at enrolment is input. The same feature extraction procedure is applied and the output will be matched against all the gallery templates using a distance/similarity function defined in the given feature space. The identification is then determined as that of the nearest neighbour. The situation is fundamentally different in the uncontrolled scenarios primarily due to obvious differences in the recording conditions and quality of devices used for enrolment and matching. Recognising faces in such a scenario, and in particular when matching small, LR degraded images against a gallery of high-resolution (HR) face images of a good size needs to incorporate some pre-processing, resolution-enhancing procedures. Therefore, image SR is deemed necessary for face recognition in these cases.

In unrestricted face recognition, SR has been used as a pre-processing method to improve image quality and obtain an HR facial image. He et al in (He & Zhang, 2010) developed an SR technique that constructs a HR face image from a sequence of LR images, or from a sufficient number of LR images with enough different sub-pixel information, to be processed using Gabor feature based recognition. The developed scheme achieves a 95% recognition rate when applied to the AR face database and the number of LR was 16 as well as the performance of this method reducing if the number of LR images decreasing or the performance keeps unchanged if the number reaches a threshold. Al-Azzeh et al. (Al-Azzeh, et al., 2008) proposed an SR method to generate super-resolved video sequences from LR video sequences; frames acquired from the enhanced video sequences are then used to train and test the per-

formance of the principal component analysis (PCA) based face recognition system. Although SR techniques for face recognition aim to improve the recognition rate after reconstructing a HR image, Hennings-Yeomans et al. (Hennings-Yeomans, et al., 2008) proposed a method to perform SR and recognition simultaneously. The authors used face features as prior information and SR prior to extracting a HR template that simultaneously fits the constraints of SR, as well as of the face features. Then, from the template, new feature vectors are computed for recognition.

The theory of CS, as explained previously in Chapter 3, demonstrates how a sampled signal can be faithfully reconstructed through $l_1$-optimization techniques, and in Chapter 4 we investigated various CS dictionaries that help to obtain a super-resolved image of reasonable quality from a LR image. In real-life, CS has many applications. It has been used to solve the problem of variation in illumination between the probe and gallery data (where the images are of the same size). In (Yang, et al., 2007) and (Ganesh, et al., 2009), the authors built a dictionary from the training images by arranging the given images in columns of a single matrix $A = [v_1, \dots, v_n] \epsilon \mathbb{R}^{m \times n}$. The represented object appears as a sparse representation of a test image with respect to the training images (i.e. dictionary) of multiple objects. Ideally, the test image $y$ of subject $i$ can be represented in terms of all of the images in the training set as $y = Ax_0$ where $x_0 = [0,..,0, \alpha_{i,1}, .., \alpha_{i,n}, 0,..,0]^T \in R^n$. The performance of this method has been evaluated on the AR database and Extended Yale B database without any illumination model to reduce the effect of lighting on the faces. Furthermore, in (Ganesh, et al., 2009) human faces are recognized from partial occlusions and the test image is expressed as a sparse linear combination of the training images plus a sparse error due to occlusion. On the other hand, CS has been used to solve the problem of image resolution based SR methods to reconstruct image with higher resolution (Yang, et al., 2008), (Yang, et al., 2010).

The rest of this chapter is devoted to testing the performance of CS based SR method applied to LR face images in terms of accuracy rates of a face recognition scheme. We shall conduct sets of experiments to compare the performance of a known face recognition scheme when we first pre-process the input LR face image by super resolving it in the spatial domain. We shall also investigate the use of CS-based super-resolution in the wavelet subbands of the LR images using different types of dictionaries.

## 5.2 Face Feature Extraction

We shall now describe the feature extraction scheme, adopted for our face recognition experiments. There is a variety of feature extraction scheme associated with face recognition (see Chapter 2). Here we shall use wavelet-based feature extraction schemes due to the multi-resolution nature of wavelet transforms, which provide a multi-streams face recognition scheme, one for each wavelet subband. Here we shall be using the simplest, and by far the most efficient wavelet filter, the Haar filter to decompose face images to level 3. The fact that most facial features are horizontal which are associated with high frequency content of LH subband (Sellahewa & Jassim, 2010), then the coefficients in the LH subband will be used as the main ingredients of face feature vector in our experiments. Z-score normalized LH subband coefficients of Haar wavelet transform at decomposition level three is used as a feature vector for recognition. We have also conducted similar experiments using feature vectors associated with the other subbands LL, HL & HH, but some of the recognition results will be presented in Appendix.

Note that the images at the enrolment are HR and their wavelet feature vectors form the Gallery. However, the images input for matching have to be first super-resolved before extracting the wavelet features for matching. In another case, super resolve the wavelet features of the input images directly for matching schema. The block diagram, below depicts the recognition scheme.

## 5.3 Databases and Experimental Protocol

Throughout this chapter, two face datasets of face images are used to evaluate the suitability of SR overcomplete dictionaries for use in face recognition from LR degraded images. The first dataset consists of images by two different quality cameras and at different distances resulting in face images that are of low resolution and/or inherently degraded. The second dataset is obtained by applying different types of image degrading procedures followed by down-sampling to good quality high resolution face images from a benchmark face database commonly used to test performance of face recognition scheme. The second dataset provides a good opportunity to test performance of our proposed CS-based face recognition scheme from synthesised LR images subjected to different models of image degradation.

1. **The UBHSD**: This is a Buckingham University recorded database of face images, which consists of 160 videos captured at four different ranges and each range contains 3-frames

for each subject. The videos were recorded in two sessions and each session includes videos captured in indoor and outdoor locations with two different cameras; a high definition (HD) and a standard definition (SD) camera (see Chapter 4 for more details). Video signals captured by digital imaging devices are generally digitized at resolution levels lower than that of still images, hence the quality of a frame extracted from a video sequences is lower than that obtained from a still imaging device. Moreover, the images captured at a distance are usually of lower resolution, which leads to low recognition accuracy, especially when the gallery is based on face images captured at short distances near the camera. This database was recorded at Buckingham University primarily to compare the performance of face recognition schemes when different video camera definition standards are used (HD cameras Vs SD cameras), see (Al-Obaydy & Sellahewa, 2011). Here we shall, test the viability of using CS-based super-resolution to compensate for the low resolution and low quality when frames are captured at different distances using both HD and SD cameras according to an evaluation protocols described below.

2. **The Synthesised Extended Yale B**: This Extended Yale-B database is a benchmark database of face images that is widely used to test the performance of face recognition schemes in the presence of extreme variation in lighting conditions (see Chapter 4 for full description). The cropped and normalized $192 \times 168$ frontal face images were captured indoor under 5 different laboratory-controlled lighting conditions relating to the direction of the source of light. The effect of these different lighting conditions relates to the amount of shadowing in the face images ranging from very mild to severe shadowing. In the latter cases, a large proportion of the face is obscured. However, the experiments in this thesis are not designed to test the performance of face recognition in terms of lighting conditions but rather in terms of LR and low quality images. The sets of images needed for our experiments will be synthesised from the collection of reasonably lit sections of the database. Therefore, we will be taking face images that have no or very mild shadowing, i.e. images from subset 1 and 2 (see Chapter 4) and subject them to a combination of different degradation functions and down sampling to simulate different models of degradation that could be associated with uncontrolled recording conditions. This is done by first degrading these "good quality" images using various blurring and degradation functions at different levels ranging from mild Gaussian blurring to severe environmental degradations. The LR degraded images are obtained by down sampling the deg-

radation images by a factor of two. We shall conduct face recognition for each level of degradation by first super-resolving these LR degraded images.

## 5.4 Experimental Results and Discussion

In this section, we will test the performance of various types of dictionary based SR for improved accuracy of a face recognition scheme. The experiments will conducted on the two image databases described above. For each database, we shall present two groups of experiments. The objective of the first group is to evaluate the performance of the wavelet face recognition scheme post the CS based SR pre-processing schemes which is used to reconstruct an HR image from a single degraded LR image. Whilst the objective of the second group of experiments is to test the performance of the SR methods for reconstructing wavelet subbands of decomposed LR face image, which are used as a feature vectors for face recognition.

### 5.4.1 Results of UBHSD Database

In this section, we present and analyse the results of applying our proposed face recognition for UBHSD database. The evaluation protocol adopted by (Al-Obaydy & Sellahewa, 2011) to test performance of the wavelet-based face recognition UBHSD database involves four configurations for each video recording scenario: Matching Outdoor, Matching Indoor, Unmatched Outdoor, and Unmatched Indoor. Since our focus is on face recognition post super resolution, our experiments are conducted only for two configurations, namely the Matching Indoor and Unmatched Indoor as described below:

- ***Matching Indoor Configuration***: have four test cases (e.g. $MI_1$, ..., $MI_4$) and for each test case $i$ ($i = 1, \dots, 4$), face images from range $R_i$ for each subject in *session 1* are selected as gallery images G, so gallery set in range $R_i$ consist of 60 images (3 images per subject). While, face images from all four ranges in both indoor and outdoor videos in *session 2* are used as probe images, so probe set consists of 480 images (24 images per subject). Moreover, both the gallery and probe images in this configuration from the same video resolution (i.e. either both are from the HD video camera or both are from the SD camera).
- ***Unmatched Indoor Configuration:*** for each test case $i$, similar to the Matching configuration, the gallery set contains face image from range $R_i$ for each subject in *ses-*

*sion 1*, while probe sets always takes images from all four ranges indoor and outdoor videos in *Session 2*. However, in this configuration, the gallery and probe images are from different video resolutions (i.e. one from HD and the other from the SD camera). In the two configurations, there is no overlap between the gallery and probe sets. Table 5.1 below shows the gallery and probe sets for these two configurations.

| Configuration | | Session 1 | | Session 2 | | | |
|---|---|---|---|---|---|---|---|
| | | HD video | SD video | HD video | | SD video | |
| | | Indoor | Indoor | Indoor | Outdoor | Indoor | Outdoor |
| HD | $MI_i$ | G, $R_i$ | | $P,R_{1-4}$ | $P,R_{1-4}$ | | |
| | $UI_i$ | G, $R_i$ | | | | $P,R_{1-4}$ | $P,R_{1-4}$ |
| SD | $MI_i$ | | G, $R_i$ | | | $P,R_{1-4}$ | $P,R_{1-4}$ |
| | $UI_i$ | | G, $R_i$ | $P,R_{1-4}$ | $P,R_{1-4}$ | | |

**Table 5.1:** The configuration for the UBHSD video database

### 5.4.1.1 Face Recognition on Super Resolved Face Images

Dictionary based SR method is used in four different combinations of gallery/probe image resolutions to super-resolve LR images of size $64 \times 64$ to their original size of $128 \times 128$ pixels for images captured at a distance. The LR faces images of this database are of different qualities and generated by down sampling the images that captured from SD camera. Note that, the degradation in these images is inherent in images captured by the SD camera at different distances and is not simulated by any degradation function. The learning dictionary in the spatial domain (LD-Sp) has been constructed using the training image set, which contains three good images per subject from subset one of the Extended Yale B database, unattached from gallery/probe images. The reason for not using the images captured by the HD video camera at the nearest distance for training the LD dictionary is to do with the fact that the HD images are used as gallery/probe images. In what follows, LR images are of size $64 \times 64$ captured by SD camera and in such cases, we always apply SR, using various dictionaries, or bi-cubic interpolation to increase resolution to $128 \times 128$. In addition, the identifiers of the CS-dictionaries are exactly as set previously, in Chapter 4. Note that when the LR images are used as gallery images, the actual gallery consists of their super-resolved versions of size$128 \times 128$. The corresponding experimental results are presented in Table 5.2, below:

| Gallery Set | Probe Set | SR methods | Capturing distance Range | | | |
|---|---|---|---|---|---|---|
| | | | $Range_1$ | $Range_2$ | $Range_3$ | $Range_4$ |
| $SD_{128}$ | $SD_{64}$ | LD-Sp | 76.25 | 68.33 | 71.25 | 69.79 |
| | | $LID_1$ | 76.04 | 68.12 | 71.25 | 69.58 |
| | | GRM | 76.04 | 68.33 | 71.45 | 69.79 |
| | | TCRM | 76.04 | 68.54 | 71.45 | 69.79 |
| | | $LID_2$ | 76.04 | 68.12 | 71.25 | 69.58 |
| | | ROM | 76.25 | 68.12 | 71.25 | 69.58 |
| | | IISR | 75.78 | 68.05 | 71.06 | 69.51 |
| | | BC | 76.66 | 68.75 | 71.66 | 69.79 |
| $SD_{64}$ | $HD_{128}$ | LD-Sp | 75.62 | 71.04 | 71.25 | 67.91 |
| | | $LID_1$ | 75.41 | 71.04 | 71.25 | 68.33 |
| | | GRM | 75.41 | 71.04 | 71.25 | 68.54 |
| | | TCRM | 75.41 | 71.04 | 71.45 | 68.54 |
| | | $LID_2$ | 75.41 | 71.04 | 71.04 | 68.33 |
| | | ROM | 75.41 | 71.04 | 71.45 | 68.33 |
| | | IISR | 75.62 | 70.83 | 71.45 | 68.33 |
| | | BC | 75.62 | 71.25 | 71.25 | 68.75 |
| $HD_{128}$ | $SD_{64}$ | LD-Sp | 68.95 | 65.62 | 72.08 | 74.58 |
| | | $LID_1$ | 68.54 | 65.62 | 71.87 | 74.58 |
| | | GRM | 68.75 | 65.41 | 72.08 | 75 |
| | | TCRM | 68.75 | 65.41 | 71.87 | 74.58 |
| | | $LID_2$ | 68.54 | 65.41 | 71.87 | 74.58 |
| | | ROM | 68.54 | 65.83 | 72.08 | 74.37 |
| | | IISR | 68.47 | 65.97 | 72.02 | 74.49 |
| | | BC | 68.75 | 66.66 | 72.50 | 75 |
| $SD_{64}$ | $SD_{64}$ | LD-Sp | 76.25 | 68.54 | 71.25 | 69.58 |
| | | $LID_1$ | 76.04 | 68.12 | 71.66 | 69.16 |
| | | GRM | 76.04 | 68.12 | 71.66 | 69.79 |
| | | TCRM | 76.04 | 68.12 | 71.66 | 70.00 |
| | | $LID_2$ | 76.25 | 67.91 | 71.66 | 69.79 |
| | | ROM | 76.04 | 68.12 | 71.66 | 69.37 |
| | | IISR | 75.78 | 67.84 | 71.60 | 69.31 |
| | | BC | 76.04 | 68.95 | 71.45 | 69.79 |

**Table 5.2:** Recognition accuracy rates (%) for the UBHSD database using different SR dictionaries to super resolve Full--face image, reported results on $LH_3$ subband.

1. The SR by GRM, TCRM, ROM, and SR by the learning dictionary LD in the spatial domain, produced almost identical accuracy rates with very marginal differences.
2. Surprisingly, in almost all distance ranges and evaluation protocols, the bi-cubic interpolation method achieves the highest accuracy and in only few cases, this highest accuracy is achieved by one or two other SR schemes. Moreover, even when an SR scheme outperforms the BC interpolation scheme at range 3, the difference is almost negligible.
3. Compared to the results obtained by Al-Obaydy & Sellahewa, 2011, shown below, one can see that in general BC-interpolation and SR based face recognition provide a

good cost effective alternative to using HD cameras. Note that in Al-Obaydy & Sellahewa experiments all images have the same size of 128×128.

| Gallery Set | Probe Set | Gallery Image Range | | | |
|---|---|---|---|---|---|
| | | $Range_1$ | $Range_2$ | $Range_3$ | $Range_4$ |
| SD | SD | 76.04 | 68.12 | 71.46 | 69.38 |
| SD | HD | 75.83 | 71.04 | 71.25 | 68.33 |
| HD | SD | 68.75 | 65.83 | 72.08 | 75.00 |
| HD | HD | 68.75 | 70.62 | 73.54 | 72.29 |

**Table 5.3:** Rank one Recognition rates by (Al-Obaydy & Sellahewa, 2011)

In order to explain the surprising observation that for all the evaluation protocols and at all the distance ranges, the BC interpolation is sufficient to achieve the best face recognition accuracy rates. We calculated the quality of the face images extracted at each distance range using images of the same face HD videos, captured at the same distance range, as the reference images. In the following Table (see Table 5.4), we present the average quality of 40 images at each distance range in terms of PSNR, contrast, and correlation measures.

| Measure Quality | Images Ranges | | | |
|---|---|---|---|---|
| | $Range_1$ | $Range_2$ | $Range_3$ | $Range_4$ |
| PSNR | 25.20 | 25.11 | 25.14 | 24.94 |
| Contrast | 0.9051 | 0.88 | 0.87 | 0.86 |
| Correlation | 0.5090 | 0.46 | 0.42 | 0.41 |

**Table 5.4:** Average quality of images captured at difference ranges by SD camera in an inside location.

All the measures have a decreasing trend in terms of distance range. The results also indicate that except for the correlation measure, the face images captured by the SD camera are somewhat mildly degraded relative to the corresponding images captured by the HD images. This may explain that the SR schemes succeed in improving the resolutions of the images and maintaining the already reasonable image quality but not better than the usual BC interpolation method. In fact, this indicates that success of CS-based SR schemes can be more apparent when the LR images are of poor quality. Indeed, this kind of observations are supported by the results of experiments conducted in Chapter 4 which confirmed that SR outperform the interpolation resizing schemes in terms of quality of output images when input LR images are somewhat severely degraded. Being a measure of similarity of the face images at different distance range to the HD images of the same subjects, the correlation values have an impact on discriminating power of face images at different distances. Therefore, we expect that the

low correlation values, at different distance range, have some impact of face accuracy rate. Consequently, this could to explain the rather relatively low and similarity of accuracy rates achieved by all schemes. This hypothesis will be revisited further, later in this chapter and the next chapter, on datasets of images subjected to a wide range of degradation.

As mentioned above, we repeated these experiments by using each of the other wavelet subbands ($LL_3$, $HL_3$ and $HH_3$) as the face feature vectors. The resulting table of accuracy are presented as an Appendix at the end of thesis (see Appendix A). In general, most of the remarks above remain more or less equally valid for these subbands. In terms of the performance of each subband, except in few cases, the recognition rates follow a well known trend known about wavelet based schemes (see e.g. (Sellahewa, 2006) and (Abboud, 2011)) of descending order of accuracy with $LH_3$ subband having the best performance, $HL_3$ subband the next best, $LL_3$ subband slightly lower, and the $HH_3$ the worst. This is normally explained by the fact that most face features are horizontal and the $LH_3$ by definition encapsulates the high frequency coefficients in the horizontal direction. The exception to this order may be associated with variation in face orientation, lighting conditions and/or distance from camera. Indeed, we can observe that the $HL_3$ subband performs better than the $LH_3$ subband at distance range 2 or 3 in some of the evaluation protocols.

A well-known practice in wavelet-based face recognition is to fuse several subband schemes at the score level for improved accuracy as demonstrated by many researchers (e.g. (Abboud, 2011), (Al-Assam, 2013), and (Jassim & Sellahewa, 2005). A sensible fusion approach in this case needs to be adaptive in terms of the evaluation protocol, image quality and/or distance range. However, such an approach will require extensive experimentation and to some extent is unlikely to be influenced by the main theme of this thesis which on the effect of CS-based super-resolution.

#### 5.4.1.2 Face Recognition on Super Resolved Wavelet Face Feature Vectors

In the above reported experiments, we followed the tradition of applying the SR methods in the spatial domain, and we applied wavelet-based face recognition schemes. Since, our face feature vectors are simply the wavelet-subbands at decomposition level 3, one wonders if applying SR methods directly in those frequency domain subbands can improve the quality of the feature vectors directly and thereby improve accuracy rates. This is also motivated by the observation that the BC interpolation is sufficient in achieving the best accuracy.

In order to investigate this idea, we need to generate a pair of dictionaries $(D_L, D_H)$ for each of the various types of CS-based SR schemes to be used to reconstruct a super-resolved version of the wavelet subband under consideration. Unfortunately, except for the image based learning schemes, the process of generating such pairs of dictionaries will produce the same pair of dictionaries generated for the spatial domain super-resolution. However, in the next set of experiments, we shall continue to test the use of these dictionaries to super-resolve the various wavelet subbands of the LR images as if they were wavelet-based dictionaries but the same spatial domain denotation. In the case of the image-based training pair of dictionaries, and for each wavelet subband we simply use the collection of all the corresponding wavelet subband of the high-resolution training face images and their LR images to train the wavelet-based pair of dictionaries that we shall refer to as the LD-WD scheme. Constructing the LD-WD pair of dictionaries is simply based on the same procedure, described in Chapter 4 that was used for the spatial domain except that the wavelet subbands are considered as the input images.

Figure 5.1 below shows the charts of accuracy rates obtained from the experiments conducted on the evaluation set of UBHSD images where we super resolved $LH_3$ subband to be used as feature vectors for recognition using the above described dictionaries including the LD-WD as well as the spatial version of the LD pair of dictionaries. For each pair of dictionaries, the super-resolution procedure is the same as the one described in chapter 4, section 4.4, but the iterative back project method is used to reconstruct the output wavelet frequency subband, and recognition is based on matching the super-resolved subband for any input LR image to that of the gallery. The identification results obtained the other reconstructed subbands (i.e. $LL_3$, $HL_3$, & $HH_3$) follow similar patterns and trends to those shown and discussed for the $LH_3$ subband. To avoid cluttering of tables and charts, these results are given in Appendix B. The experimental results in Figure 5.1 demonstrate that the accuracy rates of identification at different distances have increased when using the SR technique based on learnt-image dictionaries in the wavelet domain (LD-WD) as compared to the spatial domain LD-Sp. The maximum accuracy rate for the four different combinations of gallery/probe image resolutions reached 77.91%. However, the procedure of using the other SR-dictionaries to super-resolve the $LH_3$ subband of the LR images have all failed to improve the quality of the wavelet feature vectors, and the accuracy rate. In fact, at each distance ranges LD-WD produces superior performance in comparison to the various CS-dictionaries that were constructed in the spatial domain such as LD-Sp, $LID_1$, $LID_2$, GRM, TCRM, & ROM but used to super-

resolve the $LH_3$ subband of the LR images. Interestingly, the performances of all spatial domain based dictionaries are comparable in most cases with negligible differences.

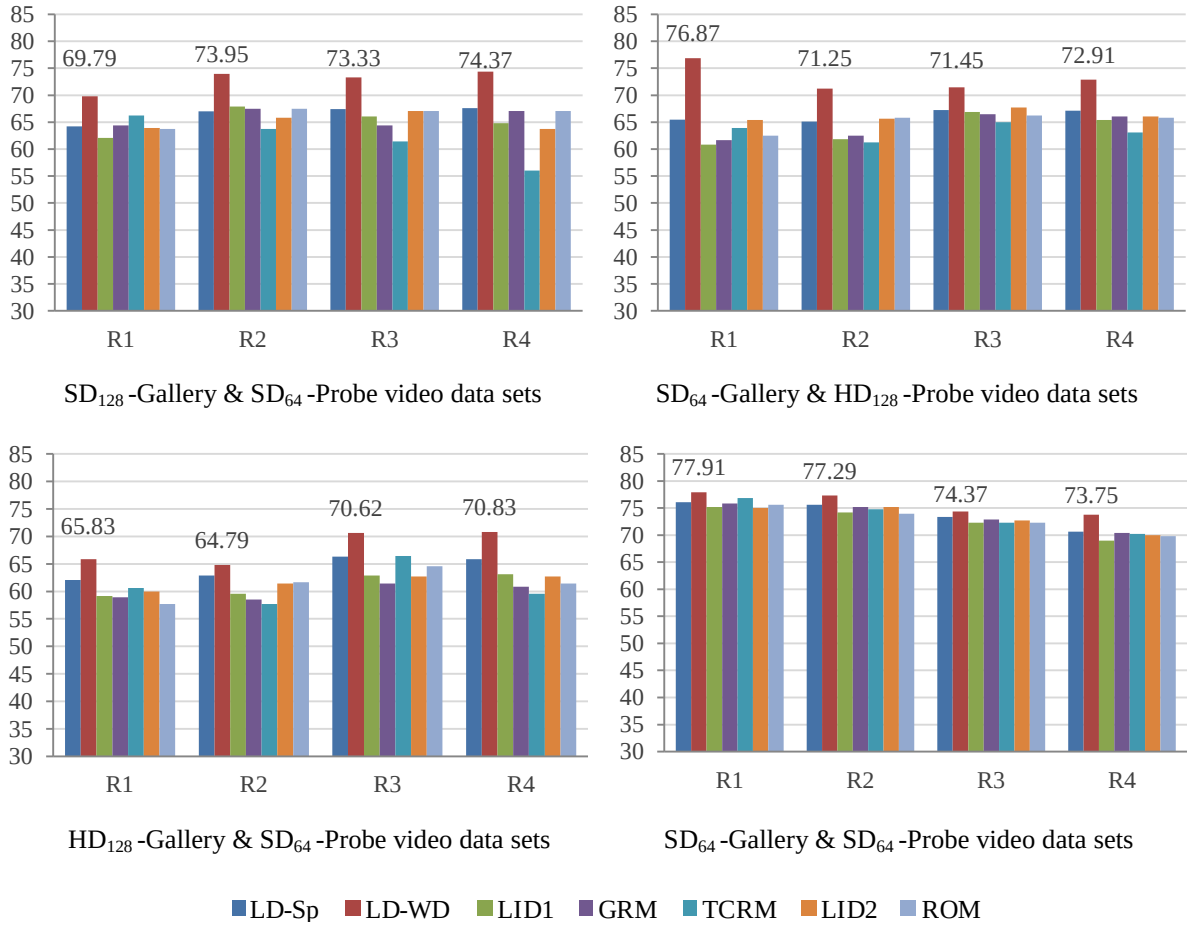



**Figure 5.1:** Comparisons between different SR dictionaries that used for super resolve $LH_3$ wavelet subband of the LR image from UBHSD data.

Comparing these results to the performance of different dictionaries that used to super resolve full-face images in the spatial domain, reported in Table 5.2 above, we observe that:

1. When the gallery consist of images captured by the SD camera, the accuracy rate achieved by super-resolving the $LH_3$ subband using the LD-WD scheme is superior to all the SR schemes that super-resolve LR images in the spatial domain, except in the first configuration at distance range 1.
2. In contrast, when the gallery consists of HD captured high-resolution images, the above trend is reversed. In fact, super-resolving the spatial domain LR images with all the dictionaries as well as the BC–interpolation yield higher accuracy rates compared

to all the schemes that super-resolve the wavelet subbands of the LR probe images including the LD-WD scheme.

Finally, the failure of the spatial domain based non-adaptive CS-dictionaries to improve the quality of wavelet feature vectors, except in the last configuration where gallery and probe images are both LR images, is most likely due to incompatibility of the dictionaries with the nature of the wavelet subband coefficients that are known to have Laplacian distributions. The success of the CS dictionaries in term of recognition accuracy when applied to super resolve the spatial domain of the LR images raises the need to design CS non-deterministic dictionaries suitable for use in the frequency domain. However, this will be considered as future work.

### 5.4.2 Results of Extended Yale B Database – Gaussian Degradation

The aim of this section is to test the effect of CS based SR methods for face recognition from frontal distorted LR still face images. For that, we need to simulate different level of blurring based distortion, applied to images in the extended Yale-B database.

We follow the same strategy adopted for the UBHSD experiments and conduct two sets of the experiments. The first set of experiments tests the performance of the various SR methods when used to reconstruct full-face image in the spatial domain. The second set of experiments tests the performance of the various SR methods when used to enhance the wavelet subbands that obtained from the input LR image. The second set of experiments includes the testing of the performance of the wavelet version of the image-learning dictionary, i.e. the LD-WD. As explained above, the LR images are generated from good quality images in subset 1 and subset 2 of the this database by first convoluting the original image by Gaussian filter with mild blur degradation of size 7 followed by down sampling by a factor of two. The Gaussian function as described in equation (4-2), in two dimensions is defined as follows:

$$\boldsymbol{G(u,v) = e^{-(u^2+v^2)/2\sigma^2}} \tag{5-1}$$

The spatial domain Gaussian filters with different sizes (e.g. $(3 \times 3)$, $(5 \times 5)$, $(7 \times 7)$ … $(15 \times 15)$) are known to have different blurring effect on the images. Increasing the size of the Gaussian filter yields increased level of blurring, and in what follows we shall test the performance of CS-based SR face recognition schemes when LR input images are blurred by a $7 \times 7$ Gaussian filter. Note that Gaussian functions in general do not reflect severe degrada-

tion that often seen in surveillance conditions, and therefore later in the chapter we consider other than Gaussian degradation models. In fact, experience from other research projects in the department has shown that the impact of increasing the size of the Gaussian filter on face recognition is rather marginal until they reach a large size whereby the image becomes useless (see (Hussein, 2014)). Therefore, we restrict our experiments in the next section to Gaussian degradation by filter of size $7 \times 7$.

### 5.4.2.1 Face Recognition on Super Resolved Face Images

We tested the accuracy rate for the various wavelet based face recognition schemes which applies CS-based SR experiments, the gallery images (one image per subject) are select from a set of face images recorded in well-lit conditions and not subject to shadowing degradation or facial expressions (i.e. subset 1), were designated as the gallery set. For testing images, the other images in these reasonably lit sets (1 and 2) will be subjected to blurring by a Gaussian filter of size $7 \times 7$, followed by down sampling by a factor of two. In total, for each degradation model we have 532 test images of size $96 \times 84$ for each recognition experiment. Table 5.5 shows the face recognition accuracy rates achieved in these experiments by each of the wavelet schemes for the subbands at level 3. The results for test images from each of the two well-lit subsets of images are given separately. Note that, the images in the two subsets differ in the direction of the light source and hence images in subset 2 may have very small amounts of shadowing while those in subset 1 have no detectible shadowing.

| Subsets | Feature Extraction | Super Resolution Methods | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | LD-Sp | $LID_1$ | GRM | TCRM | $LID_2$ | ROM | IISR | BC |
| Subset 1 | $LL_3$ | 98.68 | 98.68 | 98.68 | 98.68 | 98.68 | 100 | 95.26 | 98.68 |
| | $LH_3$ | 92.10 | 92.10 | 92.10 | 92.10 | 92.10 | 92.10 | 92.36 | 92.10 |
| | $HL_3$ | 92.10 | 92.10 | 90.78 | 92.10 | 90.78 | 92.10 | 92.10 | 92.10 |
| | $HH_3$ | 78.94 | 78.94 | 78.94 | 78.94 | 78.94 | 78.94 | 78.42 | 78.94 |
| Subset 2 | $LL_3$ | 78.72 | 78.72 | 78.72 | 78.72 | 78.72 | 79.82 | 79.16 | 77.63 |
| | $LH_3$ | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |
| | $HL_3$ | 95.61 | 95.83 | 96.27 | 95.61 | 96.05 | 100 | 95.83 | 95.83 |
| | $HH_3$ | 98.02 | 98.24 | 98.24 | 98.24 | 98.24 | 98.24 | 97.80 | 99.56 |

**Table 5.5:** Recognition accuracy rates (%) for the Extended Yale B database using different SR dictionaries to super resolve Full--face image in the spatial domain.

These results demonstrate that for each subband and each subset, there is no significant difference in identification accuracy rates between the CS learnt dictionary LD-Sp that is created using training high-resolution and high quality images, and the various types of CS-

dictionaries that constructed without using images. Moreover, the various dictionaries that are used to super-resolve the LR images can achieve high accuracy rates, which in some cases reaches 100% at least for one subband (e.g. $LH_3$ subband feature are used in subset 2 or when ROM dictionary was used in subset 1). In addition, the performance of different dictionary methods based SR in terms of face recognition produced marginally better results, if not the same, than what was achieved by the non-CS iterative SR method and/or bi-cubic method.

To compare the average accuracy achieved in these experiments by $LH_3$ recognition scheme post the use of the various SR schemes with the previous results that achieved for the UB-HSD database. We calculated the PSNR, contrast and correlation measures for 20 Gaussian blurred images in each of the two subsets, and reference images being the high quality images in the Extended Yale-B database (see the results in Table 5.6, below). In general, these results also support the previously hinted at hypothesis that when the LR images are mildly degraded simple bi-cubic method is sufficient to get excellent accuracy. Having relatively, high PSNR and contrast explain the observation that all SR schemes almost the same performance. While the relatively high correlation values (0.81 for subset 1 and 0.82 in subset 2) in comparison to the low correlation value for the UBHSD database (ranging from 0.41-0.50) explains the high accuracy rate achieved here (92% for subset 1 and 100% for subset 2).

| Subsets | Measures of Quality | | |
|---|---|---|---|
| | PSNR | Contrast | Correlation |
| Set 1 | 27.62 | 0.92 | 0.81 |
| Set 2 | 27.29 | 0.92 | 0.82 |

**Table 5.6:** Average quality values of different quality measures for 20 images in the presence of degradation.

Finally, we point out that fusion of two or more wavelet schemes would be expected to achieve the optimal accuracy.

### 5.4.2.2 Face Recognition on Super Resolved Wavelet Face Feature Vectors

We conducted another set of experiments, in similar manner to what was done in the second set of experiments on the UBHSD database, to test the impact of SR in the wavelet domain on the quality of face feature vectors of LR Gaussian blurred images and on the face recognition accuracy. Again, we note that only the genuinely wavelet-based pair of SR dictionaries is the LD-WD obtained from the subbands of the wavelet decomposed images of an election of the good quality images in subset 1. All other SR-pairs of dictionaries are generated independently of images which were used in the last set of experiments to super-resolve the spa-

tial domain LR blurred images. Table 5.7 shows the identification accuracy rates for the experiments conducted on the face images in subset 1 and subset 2 where we super resolved different subbands to be used as feature vectors for face recognition. We can see from these results that again the SR by LD-WD improved the recognition accuracy at different wavelet subbands in subset 2, and in subset 1 only when the reconstructed $HH_3$ subband is used for matching. Just as in the case of the UBHSD database experiments, the LD-WD method produced superior improvement than SR by LD-Sp as well as the other dictionaries that constructed without using images information. However, for the Extended Yale-B database the difference in accuracy rates between LD-WD and other various dictionaries that were created in the spatial domain (i.e. LD-Sp, $LID_1$, $LID_2$, ROM, GRM and TCRM) is reduced. In fact, there is no significant difference between them when $LL_3$, $LH_3$, and $HL_3$ in subset 1 are used for recognition. Again, for subset 1 and subset 2 there is little difference in identification accuracy rate between the dictionaries that constructed in the spatial domain.

| Subsets | Feature Extraction | Super Resolution Methods | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | | LD-Sp | LD-WD | $LID_1$ | GRM | TCRM | $LID_2$ | ROM |
| Subset 1 | $LL_3$ | 98.68 | 100 | 100 | 100 | 100 | 98.68 | 98.68 |
| | $LH_3$ | 97.6 | 97.36 | 98.68 | 98.68 | 97.36 | 97.36 | 98.68 |
| | $HL_3$ | 93.40 | 94.73 | 94.73 | 93.42 | 93.42 | 96.05 | 93.42 |
| | $HH_3$ | 52.36 | 77.89 | 57.89 | 52.63 | 52.63 | 53.94 | 55.26 |
| Subset 2 | $LL_3$ | 72.24 | 76.97 | 69.51 | 69.95 | 71.27 | 70.17 | 69.95 |
| | $LH_3$ | 96.12 | 99.06 | 96.92 | 96.71 | 96.49 | 96.05 | 96.49 |
| | $HL_3$ | 94.90 | 99.12 | 95.83 | 94.73 | 90.57 | 96.49 | 96.27 |
| | $HH_3$ | 65.74 | 96.49 | 69.51 | 66.88 | 59.86 | 67.32 | 70.61 |

**Table 5.7:** Comparison between different SR dictionaries that used for super resolve wavelet subbands of the LR image from Extended Yale B data.

This pattern of performance can be explained in the same way we explained the results of the experiments conducted on the UBHSD to test the impact on wavelet feature vectors.

### 5.4.3 Results of Extended Yale B Database – Turbulence Degradation

Gaussian degradation function has low level of blurring effect on any image. On the other hand the atmospheric turbulence degradation function, described in chapter 4, of different strengths has different complex degradation effect on an image that could be more reflective of the sever degradation seen in surveillance scenarios can be used as an alternative to the Gaussian model of degradation. The turbulence function as described in equation (4-3) is defined as follows:

$$\boldsymbol{H(u,v) = e^{-k_1(u^2+v^2)^{5/6}}} \quad (5\text{-}2)$$

The induced degradation level is dependent on the value of $k_1$, where different value of $k_1$results in three categories of degradation: Minor degradation when $k_1 \epsilon (0,0.02]$, Medium degradation when $k_1 \epsilon (0.02,0.04]$, and Sever degradation when $k_1 \epsilon (0.04,0.09]$. The variation in the level of blur degradation has vey affect on the contrast quality of any image and sever blur degradation could change the appearance of the face drastically. Therefore, for practical purposes we can use these values of $k_1$ and contrast quality measure for estimating the level of degradation to the surveillance image.

### 5.4.3.1 Face Recognition on Super Resolved Face Images

Super resolution based on different CS dictionaries to reconstruct a good quality image from any degraded LR image for face recognition was also tested. As described before datasets of LR face images were generated from the original images in subsets 1 and 2 after blurring with the Turbulence function using different $k_1$values and down sampling by a factor of 2. We repeated the same set of experiments to super-resolve the LR probe images for face recognition using the various SR schemes including the IISR and the BC interpolation schemes. For the sake of comparison, these experiments also include the use of matching in the low resolution, where the degraded LR images are matched against the gallery images after down sampling the gallery images by a factor of 2.

Figure 5.2 below, shows the accuracy rates for the different subbands separately for each of subset 1 and subset 2 of face images, at each level of degradation. These charts reveal the following observations:

1. Almost identical accuracy rates for all the SR dictionary methods as well as the IISR scheme.
2. For each subset, at least for one subband all the SR dictionary methods as well as the IISR scheme achieve 100% accuracy.
3. The performance of all the SR dictionary methods as well as the IISR scheme on subset 1 differs from that on subset 2. The performance of these schemes is only higher for subset 1 than for subset 2, and with a big margin, only when the $LL_3$ subband is used as the face feature vector. When the other subbands ($LH_3$, $HL_3$ or $HH_3$) these schemes attain higher accuracy on subset 2 than on subset 1. This pattern may be attributed to the combined effect of two factors: the slight differences in lighting condi-

tions between the images in the two subsets, and the quality of LL subband is affected the further the lighting source is from the straight direction. The lighting condition of subset 2 results in the appearance of some shadows around some facial features and the LL filter (also known as the approximated filter) reduces further the contrast in these regions. It is also worth noting that the gallery is constructed entirely from images in subset1.

4. Except for the $HH_3$ wavelet subband, the accuracy rates for all the SR dictionary methods as well as the IISR scheme remain the same over all degradation levels.
5. For the $HH_3$ subband, the performance of the accuracy rates for all the SR dictionary methods as well as the IISR scheme improves as the value of $k_1$increases, i.e. as degradation worsen.
6. On the other hand, matching using bi-cubic interpolation scheme or the LR images decreases when $k_1$increases beyond 0.03.

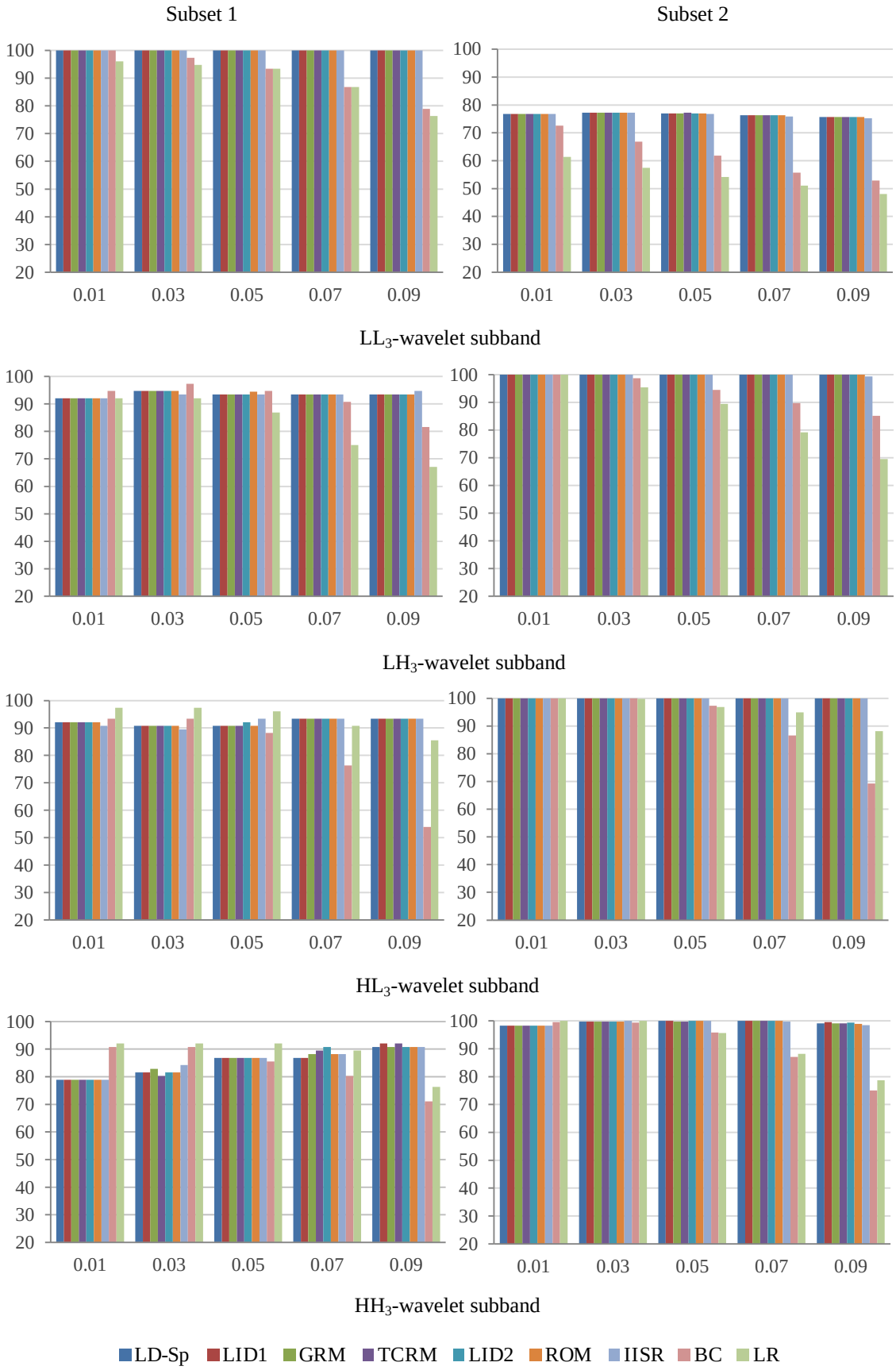



**Figure 5.2:** Recognition accuracy rates using different dictionary methods and in comparison with matching in low-resolution, bi-cubic interpolation method as well as non-dictionary iterative SR method.

In order to determine if the above pattern of accuracy of the various SR schemes as well as the BC interpolation scheme is influenced by the use wavelet-based face feature vectors, for the differently degraded LR face images. We conducted face recognition with exactly the same degraded LR images but using Principal Component Analysis (PCA) on the spatial domain face feature vectors. PCA is usually used to reduce the dimensionality of the spatial domain face data, simply using the eigenvalues corresponding to the most-significant eigenvalues of the covariance matrix of the face images in the training set (Turk & Pentland, 1991). In our experiments, the face images in the gallery set has been used for training the PCA and the output feature vector after applying PCA technique is of size 37.

The results of the conducted experiments are shown in Figure 5.3, and demonstrate that across different degradation functions, the pattern of accuracy rates, for the different testes schemes, is similar to those shown above for the $LL_3$ subband. However, the difference between the accuracy rates for subsets 1 and 2 is less significant than in the previous Figure (i.e. Figure 5.2). For both subsets the performance of the PCA recognition scheme post the bicubic interpolation method, deteriorates significantly as the $k_1$degradation value increases. The same is true when the PCA is applied to the LR images. On the other hand, the SR methods provide significant improvements in performance of the PCA face recognition system even when the degradation is severe. Consequently, we expect that the pattern of accuracy of face recognition post SR schemes is to some extent is independent of the face feature extraction scheme.

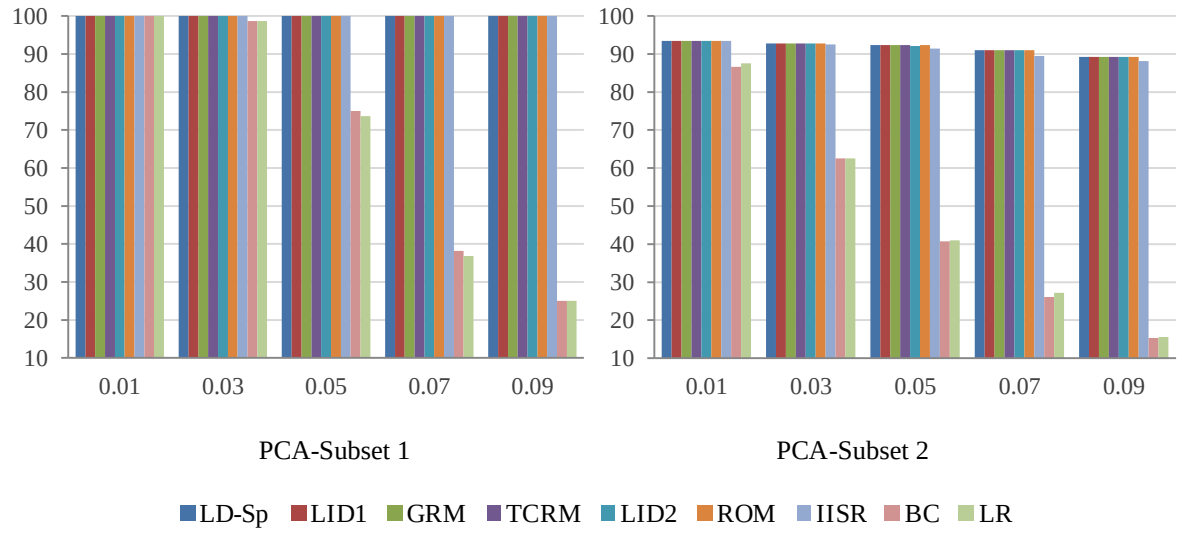



**Figure 5.3:** Recognition accuracy rates (%) for the Extended Yale B database based on PCA.

In summary, various CS dictionaries based on SR methods can improve the performance of face recognition for images with low-resolution and low quality. The non-adaptive dictionaries that are constructed without using image information can achieve identification accuracy rates that are good and in some cases slightly better, if not equal to, the learning dictionary based on the training images in the spatial domain. Moreover, the efficiency of the SR based different dictionary techniques are not affected by different degradation levels while maintaining the accuracy rates at the same level.

#### 5.4.3.2 Face Recognition on Super Resolved Wavelet Face Feature Vectors

In this section we present the results of experiments we conducted on the differently degraded and down-sampled Extended Yale-B images, in a similar manner to those conducted in the earlier section for the UBHSD database that tests the effect of super-resolving the wavelet subband rather than the spatial domain images. Various CS dictionary methods based SR that are constructed in the spatial and wavelet domains to reconstruct the feature vectors ($LL_3$, $LH_3$, $HL_3$, & $HH_3$) of any degraded LR images from the Extended Yale B database was tested for face recognition. The results shown below in Figure 5.4, and demonstrate similar patterns observed in the experiments conducted on the UBHSD database (Figure 5.1). We can see from Figure 5.4 that the accuracy rate achieved by the various SR dictionaries including the LD-DW wavelet domain training dictionary. Here we are not including the performance when using the bi-cubic interpolation or using LR matching with gallery being down-sampled. From Figure 5.4 we note that at each level of degradation and for each wavelet subbands from the two subsets, the SR method by LD-WD outperforms the scheme that uses the LD-Sp as well as the various types of non-adaptive spatial domain dictionaries (i.e. not learnt using large sets of images wavelet subbands). Moreover, except for the $LL_3$ and $HH_3$ subband feature vectors, increased level of degradation yields decreased identification accuracy rates by almost all dictionaries (including the LD-WD). In the $LL_3$, and to less extent in the $HH_3$, accuracy rates remain more or less the same at different degradation level.

As in the case of the similar experiments on UBHSD database, these results demonstrate the success of the LD-WD scheme, at all degradation levels, in improving the quality of the super-resolved subbands, more than what was achieved by the other SR dictionaries. However, in this case the difference is somewhat less significant than the case of the UBHSD database. Again, we believe that this pattern of quality enhancement of the wavelet feature vectors is most likely due to incompatibility of the dictionaries with the nature of the wavelet subband

coefficients that are known to have Laplacian distributions. This observed success, albeit marginal, of the CS dictionaries in term of recognition accuracy when applied to super resolve the spatial domain of the LR images re-enforces our earlier remark about the need to design CS non-deterministic dictionaries for use in the frequency domain as a future work.

Finally, we observe that the performance of the various schemes differ significantly between the two subsets for the $LL_3$ feature vectors. Again, this is most likely due to the variation in lighting condition between the two subsets.

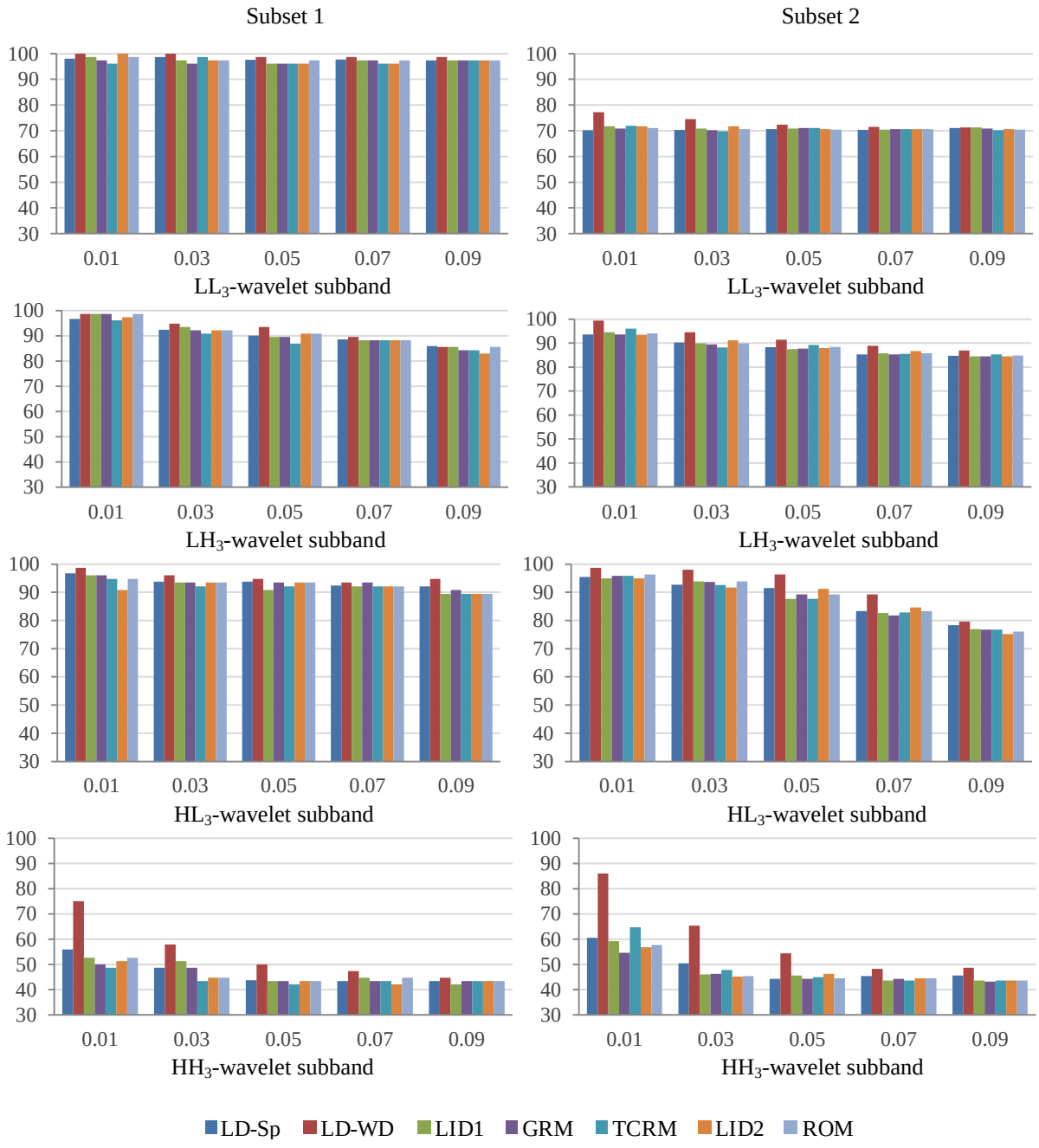



**Figure 5.4:** Comparisons between the various SR dictionary in the spatial / and wavelet domain to super resolved wavelet subbands of the degraded LR image from the Extended Yale B data.

## 5.5 Summary and Conclusion

In this chapter, we investigated the effect of image super-resolution on face recognition where the capture face images may be subject to low resolution as well as different level of image degradation. In general, the biometric community agree that face recognition from quality degraded low-resolution images, associated with recording in uncontrolled environments, is a tough challenge that needs to be tackled due to increased interest in security and safety applications. Here we considered two possible ways that such scenario may arise: recognition of faces from video frames recorded by different quality cameras captured at different distances, and recognition from simulated low quality low resolution images obtained from high quality and resolution images by applying different models of degradation followed by down-sampling. Different sets of experiments were conducted to test the performance of a variety super resolution schemes for dealing with those two cases. The first set of experiments, aimed at comparing the use of super-resolution instead of using High definition video cameras to overcome the problem of face recognition at 4 different ranges. The performance of HD camera have been shown by Al-Obaydy & Sellahewa (Al-Obaydy & Sellahewa, 2011) not to provide added benefits in terms of recognition accuracy at close range distances.

We investigated the use of various overcomplete CS dictionaries such as learnt dictionaries in the spatial or in the wavelet domain using sufficiently large training set of face images, non-adaptive image-independent random or implicitly designed dictionaries that are known to "satisfy" the RIP condition.

The results of the first set of experiments demonstrate that Super-resolving the wavelet $LH_3$ subband of LR image captured by SD camera, using the overcomplete wavelet based LD-WD dictionary that was learnt from the $LH_3$ subbands of a selected set of good quality training images does certainly obviate the need for costly HD cameras for face recognition at a distance. Moreover super resolving the LR images using the spatial CS-based dictionaries, including the LD dictionary, have similar performance, which is better if not equal to the HD cameras for recognition but with a slightly lower accuracy to that achieved by the LD-WD dictionary. These conclusions are based on comparing the results with those obtained by Al-Obaydy & Sellahewa, 2011. All these results are based on using a single wavelet subbands, and significantly, higher accuracy can be achieved by fusing at the score level two or more subbands.

The video frames used in the first set of experiments, are not severely degraded, and hence we cannot be assured that the SR approaches can be used to deal the challenge of face recognition when the images are severely degraded besides being of low resolutions, such as the images captured by CCTV cameras. Therefore the second set of experiments, were designed to test the performance of wavelet-based and PCA-based face recognition schemes post SR procedures, using various dictionaries, on a simulated data set of degraded LR images obtained from two subsets of the Extended Yale-B database that only differ slightly in lighting conditions. The simulated degradation was obtained by two models of blurring (Gaussian and Environmental distortion) with different input parameters to control the severity of the degradation. The conclusions from this set of experiments have resulted in a similar performance patterns to those obtained from the first set of experiments, but with higher accuracy rates achieved by single wavelet subbands. In particular, these experiments have confirmed that CS-based super-resolution using non-adaptive dictionaries, as well as learnt dictionaries both in the spatial and wavelet domain, help mitigating the adverse effect of a wide range of image degradations on face recognition, especially in the case of severe image degradations. In fact, these accuracy levels are hardly affected by different degradation levels and maintained the accuracy rates at the same level. The results of using PCA instead of wavelet based recognition schemes have illustrated that the performance of CS-compliant dictionaries is independent of the feature extraction schemes.

More importantly, in both sets of experiments the non-adaptive dictionaries that were constructed without using image information can achieve face recognition accuracy levels as good as, if not better than, those achieved by dictionaries that depend on the spatial domain training set by super resolved image from the LR image in the spatial domain. These results remove the need of training face images for recognition purposes. However, the rather superior performance of the learned dictionary LD-WD in super-resolving the wavelet subbands of the LR images, in comparison to the other non-adaptive dictionaries, reveal that the current approaches to the construction of non-adaptive dictionaries are not suitable to super-resolve every type of signals and certainly improve the quality of wavelet subbands of images. This may be due to the incompatibility of the non-adaptive dictionaries with the characteristics of wavelet coefficients that are known to have Laplacian distributions. Moreover, the success of the CS non-adaptive dictionaries in term of recognition accuracy when used to super-resolve the spatial domain of LR images raises the need to investigate new design strategies for the construction of CS non-adaptive dictionaries suitable for use in the frequency domain. Although, this is outside the objectives of this thesis, it will be part of the follow up future work.

Finally, the various simulated degradation models, including the severe ones, do not reflect the scenario of CCTV cameras, because the quality of images are effected by a complex set of factors including variation in lighting/environmental conditions, pose, and movement of subject relative to the camera and sources of light as well as camera quality. However, the results from both sets of experiments is a strong motivation to investigate the use SR based techniques that employ CS dictionaries for recovering images from a database of genuine CCTV face images with different resolution cameras. This will be the main objective of the next chapter.

# Chapter 6

## COMPRESSIVE SENSING & SUPER RESOLUTION FOR CCTV FACE IMAGES

In the previous two chapters, we have shown that compressive sensing based super resolution methods help to improve image quality and image resolution as well as face recognition rates. However, in those chapters we were dealing with frontal LR face images whereby the model of image degradation were based on different levels of Gaussian blurring as well as environmental image parameterized frequency domain degradation exponential filters. In fact, for these experiments the LR face images were generated from good quality images obtained from a benchmark face image database by applying different levels of Gaussian/environmental blurring/degradation functions followed by downsampling. However, degradation of CCTV images is difficult to model due to a variety of complex combinations of factors influencing the process of capturing face images in ways that are highly unlikely to be modelled by simple exponential domain frequency functions. Moreover, CCTV surveillance video images generally include faces with varying poses, illumination and different resolution due to the distances between the camera and the photographed person(s). Face images captured at a distance from the CCTV suffer from reduced image resolution, which results in loss of high frequency facial components with adverse impact on recognition rates. Therefore, identifying an individual from CCTV camera remains an extremely challenging problem that significantly accedes the challenges dealt with in the previous chapters. In this chapter, we shall nevertheless investigate the use of overcomplete SR dictionaries that satisfy compressive sensing properties to restore surveillance images and test the performance of wavelet-based face recognition using the super-resolved face images. We shall first attempt to approximate the model of CCTV image degradation by parameterized exponential frequency domain environmental distortion functions, and demonstrate modest accuracy rates of face recognition that are significantly higher than the state of the art schemes. For testing purposes, we simulate a very challenging recognition scenario by using a database of 130 moving sub-

jects where the face images/or video is recorded under different resolution, illumination, expression and in varying poses.

## 6.1 Introduction

Face recognition has been investigated extensively in recent decades and has many applications, such as passports and driving license cards. Most, face recognition algorithms have reasonable recognition accuracy in controlled environments, but their accuracy drops dramatically in uncontrolled environments. Much research in the area of face recognition has been concentrated on recognizing faces under one challenge. such as across changes in illumination (Xie, et al., 2008), (Jassim & Sellahewa, 2005), (Vuini, et al., 2007) or across different poses (Prince, et al., 2008), (Zhang & Gao, 2009), (Shiau, et al., 2010) while little has been done for dealing with combined effect of several challenges (Peng, 2011), (Biswas, et al., 2011). The case of combined challenges is more evident when using CCTV cameras. In fact, the quality of images captured from the CCTV video camera in uncontrolled conditions is much lower than that obtained from that of still imaging. With increasing demands for surveillance camera-based applications, the need to recognize faces captured on such devices, which are mostly of low-resolution in uncontrolled poses/illumination, is in demand. In this chapter, we are concerned with face recognition where the face images are captured by CCTV cameras.

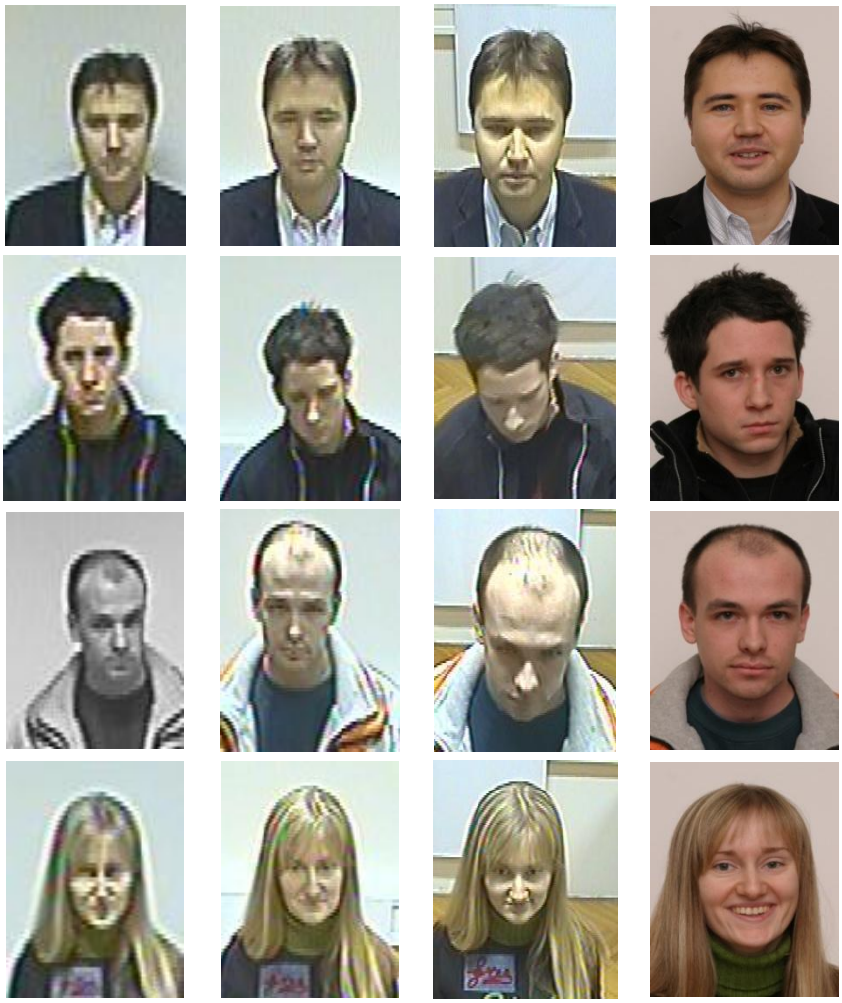
Face recognition systems generally perform well when the gallery and probe images have the same resolution and size as well as taken under similar controlled imaging conditions. However, differences in information content between LR images (obtained from surveillance cameras) and HR images (i.e. images captured during enrolment) greatly degrades the performance of existing face recognition algorithms. Two approaches have been adopted to address this challenge; match cross resolution images by downsampling HR images to the level/ size of LR images before matching; or alternatively use the SR technique to enhance the LR face image before matching. However, in the first case the information needed for face recognition (such as texture, edges and other high frequency information) is compromised in the process of downsampling the images. In the second approach (Al-Azzeh, et al., 2008) the objective of the SR method is to construct high frequency details that are missing in the LR image.

In the earlier chapters, we discussed different SR schemes, including compressive sensing based schemes that uses learnt as well as non-adaptive over-complete dictionaries. (Peng, 2011) and (Zou & Yuen, 2010) proposed SR algorithm to reconstruct an HR image for face recognition by training a dataset of HR and LR image pairs ($\{I_i^h, I_i^l\}_{i=1}^N$) to learn first the relationship/or the mapping $R$ between the pairs of images; then the SR image is reconstructed by applying R on the testing LR image. The face LR images are generated from the surveillance face data.

In the rest of the chapter, we test the performance of our compressive sensing based SR approaches using the previously investigated CS-relevant dictionaries to super resolve face image from surveillance images for face recognition. For comparison, we shall also test the performance of the standard interpolation method.

## 6.2 Experimental Database and Testing Protocol

In order to evaluate the performance of the various overcomplete CS dictionaries based SR method for face recognition on the surveillance images with different resolutions; the publicly available surveillance cameras face (SCface) database (Grgic, et al., 2011) will be used, where various quality and resolution cameras have. SCface database is a set of static face images, which contains 130 subjects from both genders, recorded in an uncontrolled indoor environment using five different quality surveillance video cameras, namely cam1, cam2, cam3, cam4 and cam5. A high quality camera is also used to capture an HR frontal image for each individual in indoor controlled conditions. Moreover, this database was collected with the cameras placed slightly above the subject's head and during the recordings; the subject was not required to look at a fixed point. In each recording session, the image of each subject was captured at three different distances from the cameras, close (distance 3), mid (distance 2) and far (distance 1) of 1.00, 2.60 and 4.20 meters respectively. Hence, the total number of images in the database is 4,160 images, varying in resolution, pose, illumination and expression. Figure 6.1 shows sample images from the SCface database that captured by cam1 at the three different distances. Although the images from a distance 3 are near to the camera and with higher resolution than images at distances two and one, the pose variation is more serious.



(a) (b) (c) (d)

**Figure 6.1:** Images from SCface database. (a) The first column images from distance 1. (b) Second column: images from distance 2. (c) Third column: images from distance 3. (d) Last column: HR frontal images.

All images in the SCface database are originally coloured, but for our experiments are transformed into grey-scale images, the faces are cropped and resized to $64 \times 64$ pixels. Figure 6.2 shows the cropped and rescaled face images for the same person in the first row in Figure 6.1 on five different quality surveillance cameras.

Super resolution techniques, in the previous chapters, have been used to improve the quality of under-sampling LR images by increasing their spatial resolution and attempting to filter out image degradation such as blurring (or environmental distortion) at different levels. The success of this step depends on some knowledge/estimation of the degradation present in the probe CCTV images. This is the first challenge that we shall deal with.

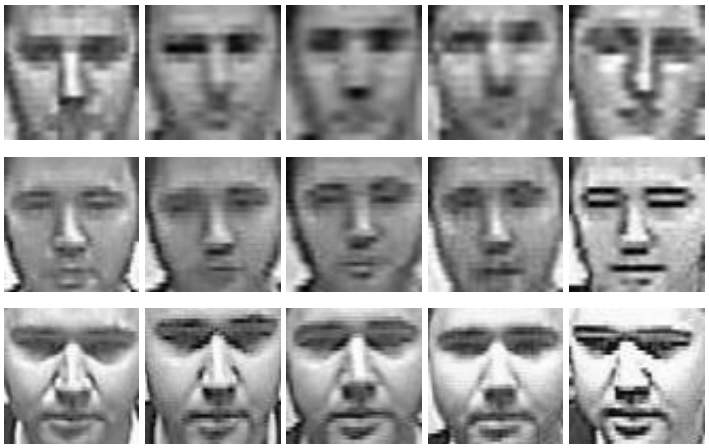



**Figure 6.2:** Examples of cropped and rescaled face images from SCface database. The first row images from distance 1; second row: images from distance 2; third row: images from distance 3.

## 6.3 Towards a Model of CCTV Image Degradation

The image quality is significantly affected by a number of factors related to the recording conditions, including environmental factors, camera specification, the orientation of the camera relative to the photographed object, and distances between from the photographed object. The modelling of image degradation for surveillance images at any distance, such as the images in the SCface database captured by the five different cameras, depends on a number factors including camera quality and environmental factors. Knowledge of good model of degradation is essential for super-resolution. In fact, all SR techniques use an inverse filtering at some stage to improve the quality of super-resolved images, the appropriate filter must be based on estimation of the degradation transform. A possible solution could be based on conducting an extensive set of experiments to figure out the structure of the frequency spectrum of a large set of CCTV images in order to estimate a parameterized frequency domain function that models different levels of image degradation of CCTV video frames.

Below, we display the Fourier spectrums for some images from SCface database, captured by different cameras at different distances. Figure 6.3 and Figure 6.4 illustrate that lightening condition, distance, and camera type have varying effect on the Fourier spectrum of the images. Note the distortion around the u and v axes compared to the Fourier spectrum of the original HR image. It is clear that the distortion around the vertical axis in particular decreases as the distance to the camera decreases. A more systematic investigation into this will, and should, be the subject of future work.

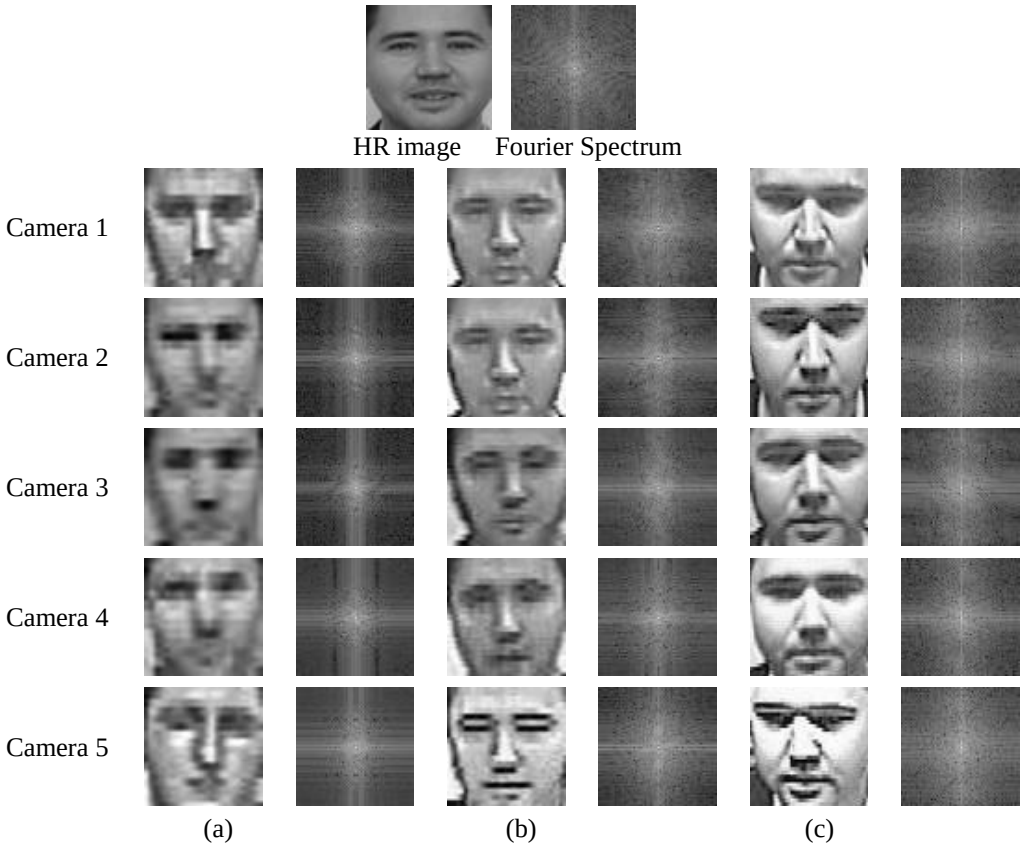



**Figure 6.3:** Example 1, Fourier Spectrum for: HR image from SCface database and for each image captured for the same person by different cameras from (a) distance 1 (b) distance 2 (c) distance 3.

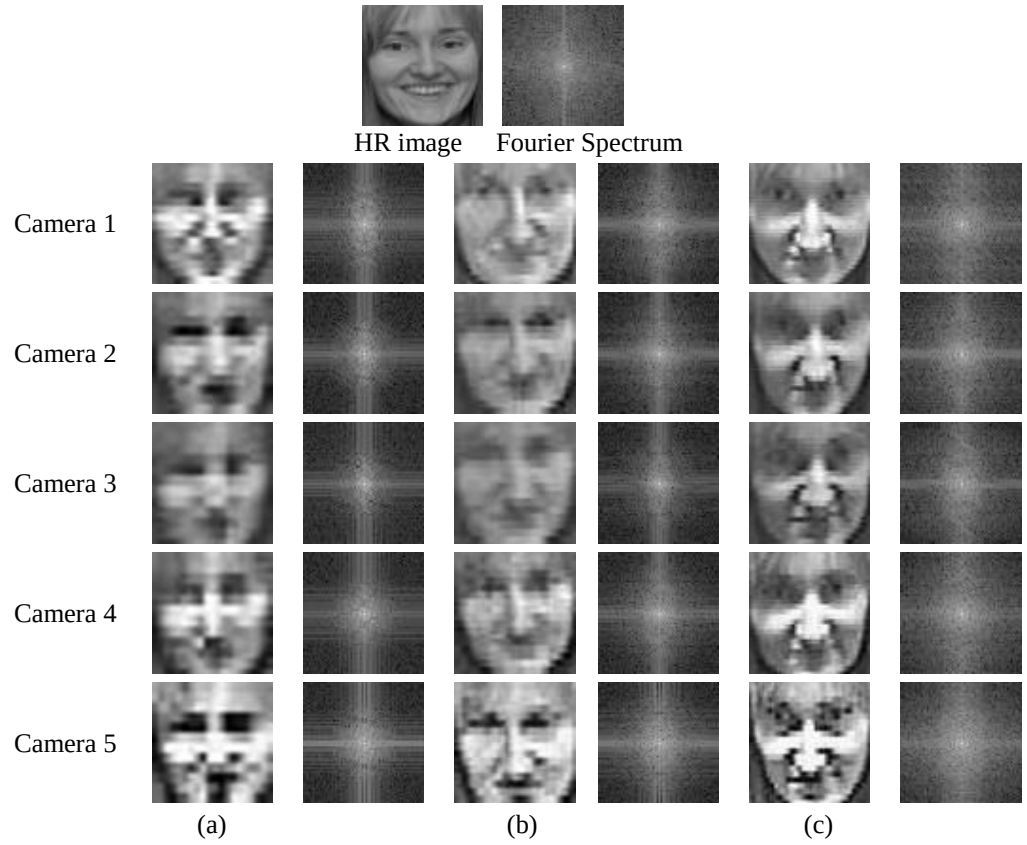



**Figure 6.4:** Example 2, Fourier Spectrum for: HR image from SCface database and for each image captured for the same person by different cameras from (a) distance 1 (b) distance 2 (c) distance 3.

Conducting the necessary investigation to determine the frequency domain degradation filter require extensive and challenging work which, to large extent, is outside the scope of this thesis but would be the subject of future work. Hence, we need a simpler alternative strategy.

When mathematicians are faced with such difficult problems either one reduce the complexity of the problem by solving a less challenging provisions of the problem or attempt to approximate the required model of degradation by known, albeit different, models. Hence, we shall use an indirect method to link/approximate the CCTV degradation function to/by a known degradation function through a common, and easy to compute, image quality measure. In fact, the work done in the last two chapters is reasonably sufficient for our work here. Therefore, in this chapter, we adopt a general approach to approximate the required degradation functions using an existing simple, and easy to compute, image quality measure that has "monotonic" relationship with the environmental degradation function, studied in the last chapter. Ideally, the monotonic relationship remains valid for CCTV images, so that one can then approximate the unknown CCTV degradation model by linking to an appropriate $k_1$value for environmental degradation.

First, we observe that most images in the SCface database have poor contrast and it would be natural to measure their contrast value with reference to the provided high quality images in the database for the same persons. Hence, it is natural to compare the contrast value of an input LR image (relative to the HR face images) with the contrast that can be obtained when different ranges of $k_1$-values are used to degrade HR images. Another, related image quality measure that could be investigated for linking environmental degradation functions in terms of $k_1$-values is that of correlation. Here we recall that the contrast and correlation values can be obtained using the following formulae deduced from the universal image quality index (see chapter 1, section (1.4)):

$$\textbf{Correlation} = \frac{\sigma_{xy}}{\sigma_x \sigma_y} \quad \text{and} \quad \textbf{Contrast} = \frac{2\sigma_x \sigma_y}{\sigma_x^2 + \sigma_y^2}$$

Where, $\boldsymbol{x} = \{\boldsymbol{x_i} | \boldsymbol{i} = \boldsymbol{1, 2, \dots, N}\}$ and $\boldsymbol{y} = \{\boldsymbol{y_i} | \ \boldsymbol{i} = \boldsymbol{1, 2, .., N}\}$ be the column vectors representing the pixel values of reference image $x$ and an input image $y$, respectively,

$$\boldsymbol{\sigma_{xy}} = \frac{\boldsymbol{1}}{\boldsymbol{N-1}} \sum_{\boldsymbol{i=1}}^{\boldsymbol{N}} (\boldsymbol{x_i} - \bar{\boldsymbol{x}})\,(\boldsymbol{y_i} - \bar{\boldsymbol{y}}),\ \boldsymbol{\sigma_x} = \left[\frac{\boldsymbol{1}}{\boldsymbol{N-1}} \sum_{\boldsymbol{i=1}}^{\boldsymbol{N}} (\boldsymbol{x_i} - \bar{\boldsymbol{x}})^{\boldsymbol{2}}\right]^{\boldsymbol{1/2}},\ \boldsymbol{\sigma_x^2} = \frac{\boldsymbol{1}}{\boldsymbol{N-1}} \sum_{\boldsymbol{i=1}}^{\boldsymbol{N}} (\boldsymbol{x_i} - \bar{\boldsymbol{x}})^{\boldsymbol{2}},$$

$$\boldsymbol{\sigma_y} = \left[\frac{\boldsymbol{1}}{\boldsymbol{N-1}} \sum_{\boldsymbol{i=1}}^{\boldsymbol{N}} (\boldsymbol{y_i} - \bar{\boldsymbol{y}})^{\boldsymbol{2}}\right]^{\boldsymbol{1/2}},\ \ \boldsymbol{\sigma_y^2} = \frac{\boldsymbol{1}}{\boldsymbol{N-1}} \sum_{\boldsymbol{i=1}}^{\boldsymbol{N}} (\boldsymbol{y_i} - \bar{\boldsymbol{y}})^{\boldsymbol{2}}.$$

and $\bar{x}$ & $\bar{y}$ are the mean of image $x$ and image $y$ respectively.

To study the relationship between $k_1$-distortions and contrast/or correlation quality measure, 20-HR images were selected randomly from the SCface database which were captured by good quality camera, degrade using increasing $k_1$-Level of environmental turbulence function, and then computes the corresponding contrast and correlation values using the original image as a reference. Figure 6.5 below show the chart obtained from this experiment. It can be seen that the image quality, based on both measures, is decreasing with an increased level of $k_1$-value distortions.

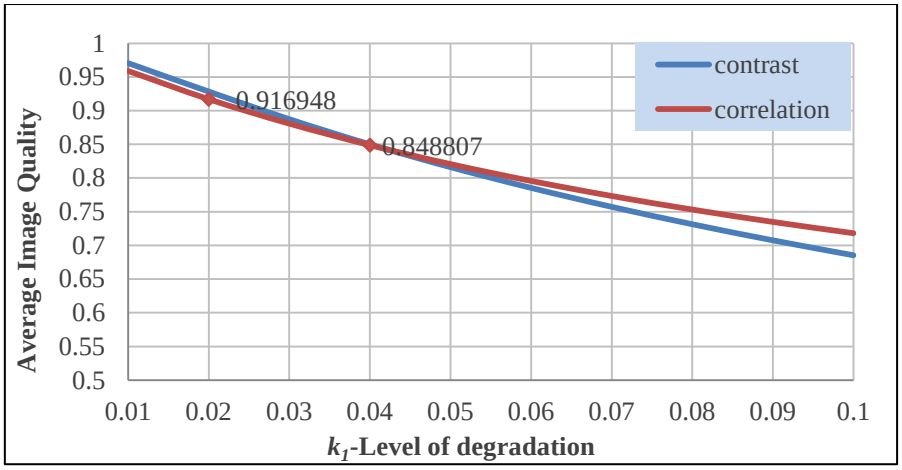



**Figure 6.5:** Contrast and Correlation measures Vs. $k_1$-Level degradation - Images from the Scface database

We repeated the same experiment, but this time we used 20 face images that were selected randomly from set 1 of the extended Yale B database, degrade by different $k_1$-levels of degradation and down-sample by factor two. Figure 6.6, shows the corresponding chart for this second experiment.

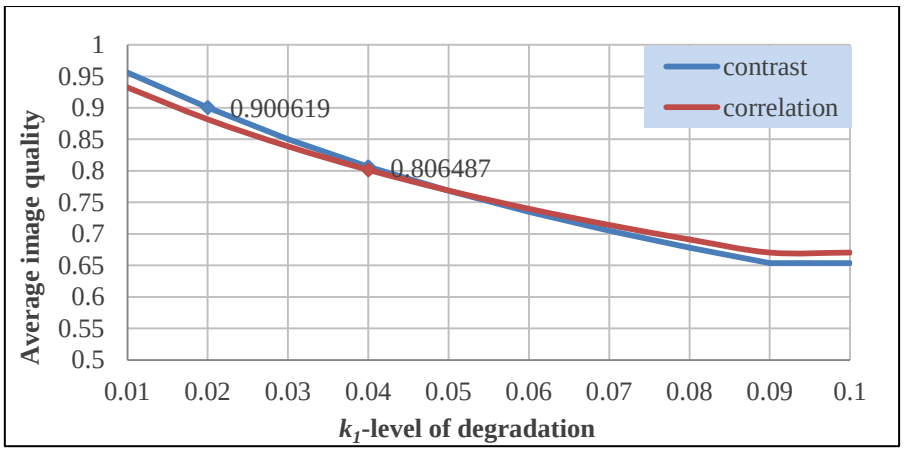



**Figure 6.6:** Contrast and Correlation measures Vs. $k_1$-Level degradation – Images from the Extended Yale B.

This experiment reveals a similar relationship to that obtained using images from the SCface database. In both cases, contrast and correlation are almost affected similarly by any $k_1$-degradation, and the only difference is that in the case of the SCface database the decrease in the contrast/correlation is slower as the $k_1$-value increases. This could be because the HR camera used in SCface database is of higher quality than that was used for recording the Extended Yale-B. We conclude that it is more sensible to use chart 6.6 when approximating the CCTV degradation model in terms of the $k_1$-value of the environmental degradation function.

For any LR image, we super resolve first, but before inverting the degradation filter, we calculate the contrast and use chart 6.6 to deduce the $k_1$-degradation level. Accordingly, we can deduce that:

1. Low $k_1$-degradation level in the range $[0, 0.02]$ corresponds to high contrast quality in the range $>= 0.9$,
2. Mild $k_1$-degradation level in the range $(0.02, 0.04]$ corresponds to average contrast quality in the range $[0.8, 0.9)$ and
3. Severe $k_1$-degradation level in the range $(0.04, 0.1]$ corresponds to low contrast quality in the range $< 0.8$,

The above investigation assumes the availability of an HR image for every person photographed by the CCTV camera. However, in the real world, the HR face image may not be available and we need to estimate the degradation level of an input CCTV image without a reference image. In this case, the Mean square error (MSE) can be used as a measure of quality to assess the quality of the LR image. This could be done by finding the error between the input image and the convolution of the input LR image by blurring filter, e.g. median or adaptive median filter, acting as its own reference image. However, this will not be pursued any further. In the next section, we shall use the contrast measure to estimate the $k_1$-value corresponding to the estimated frequency domain degradation present in any input probe image. The estimated $k_1$-value will then be input to each of the super-resolution procedures, with different dictionaries, to design the inverse frequency domain filter to improve the quality of the probe super-resolved image. In section 6.5 we shall present tables to illustrate the improvement in the values of contrast, correlation as well as PSNR (which is related to the MSE value) parameters post the SR step.

## 6.4 Face Recognition Experiments for the SCface Database

Our recognition experiments on the SCface database are conducted for each of the three image sets corresponding to the three distances (close, mid and far). To test the performance of images depending on an assumption one makes in terms of image quality/resolution, distances and camera. Here, we present two experimental configurations:

1. The first configuration: one HR frontal image per subject are selected as reference images whose extracted feature vectors from the gallery set, while the feature vectors of five images corresponding to five different cameras per distance for each subject are used for matching
2. The second configuration: five images taken by cameras 1 to 5 for 130 persons at distance 2 or at distance 3 are used as gallery images, while the probe set always takes images from distance 1, which are of lower resolution.

The main step in any biometric system is the feature extraction and matching. For this, we follow the same feature extraction scheme adopted throughout the thesis:

1. Use Haar wavelet to decompose the incoming face image to level 3,
2. Select any subband and normalise its coefficient using the standardised z-score normalisation to form the corresponding face feature vector, and
3. Use city block distance for matching.

In both recognition configurations, we shall use super-resolution to make the gallery and the probe having the same (higher resolution). In all the experiments, the resolution of gallery image is set to $128 \times 128$ pixels. Figure 6.7 and Figures 6.9 & 6.10 respectively illustrate identification accuracy results for the two gallery-probe configurations listed above.

### 6.4.1 Face Recognition Experiments – Configuration 1

For matching an LR probe with an HR gallery image, the SR technique has been used to obtain images with higher resolution than the input LR probe image, which is then used for matching with the HR gallery. Various CS dictionaries based SR methods as well as a standard bi-cubic interpolation method has been used on sets of probe images with different resolution to reconstruct HR images. In general, the facial images of SCface have very challenging illumination, pose variation and also LR, which leads to the very low benchmark performance (Grgic, et al., 2011), (Peng, 2011). Where the experimental results by (Grgic, et al.,

2011) showed that when the frontal images were used as a gallery, which of the same size of the input probe image, the results are quite low, ranging from below 1% to about 8% as well as the distance from the camera has an effect on the recognition performance. While, the results by (Peng, 2011) indicated that when the images from distance 2 or 3 used as gallery images and the super resolved images from distance 1 used as probe images, the identification rate of distance 3 is much lower than distance 2 and the improved results in distance 2 are limited which are no more than 15%.

Figure 6.7 below, shows the recognition accuracy results for different subbands at level 3 of wavelet decomposition. The various charts display rank-one recognition accuracy rates at each distance for the 4 level 3 subbands $LL_3$, $LH_3$, $HL_3$ and $HH_3$. As can be seen, the distance from the camera has some influence on the performance of face recognition where the accuracy rates at distance 3 are lower than other two distances. Note that images taken at distance 3 capture more of the top part of the head, thus covering the parts of the face normally seen in frontal gallery images and include many face features. The results in Figure 6.7 show that there is a small difference in accuracy rates between the various CS dictionary methods at each distance. However, the proposed $LID_1$ dictionary (see Chapter 4) in distance 1 yields better accuracy rates than the bi-cubic method and other dictionaries that do not depend on images when $LL_3$, $HL_3$ and $HH_3$ wavelet subbands features are used. Moreover, the performance of the $LID_1$ is far superior to LD-Sp dictionary based on images when we use $LL_3$, $LH_3$, $HL_3$ & $HH_3$ subbands as feature vectors for images in distance 1.

Although, the modest improvement of LR surveillance images for recognition is rather noticeable, the results in general are still disappointingly low. However, these results illustrate that dictionary based SR methods can achieve higher results than the recognition accuracy rate without using the SR method to the LR (Grgic, et al., 2011) where the results have never exceeded 10% recognition rates.

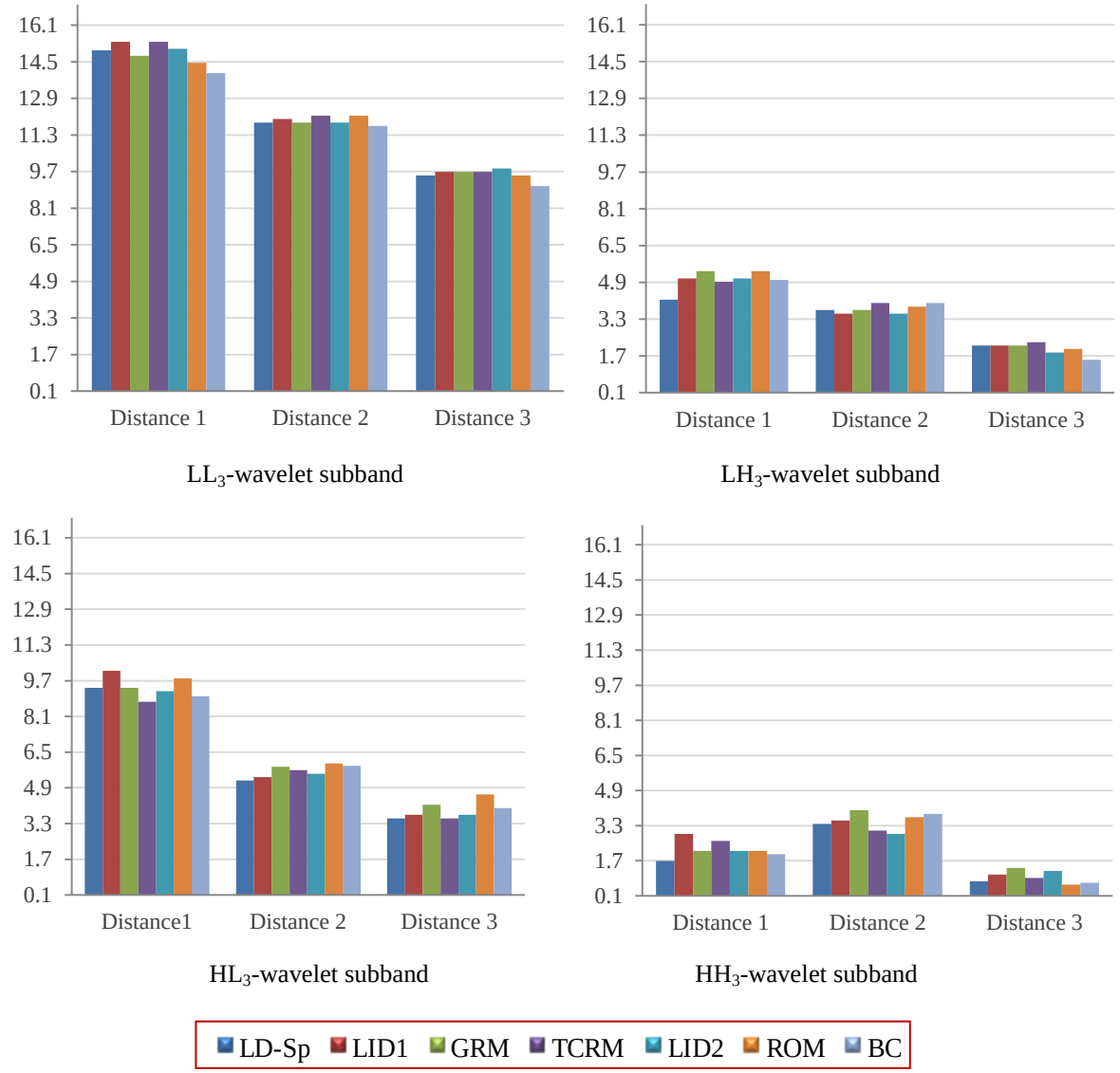



**Figure 6.7:** Rank-one recognition accuracy rates for SR images from different resolution enhancement methods.

Comparing the performance when using different subbands we, note that the best performance is achieved when we use the $LL_3$ feature vector scheme, followed by the $HL_3$ scheme, which perform better than the other two subbands. This is somewhat different from the results achieved by the investigations reported by the various publications on wavelet-based schemes (Sellahewa & Jassim, 2008), (Al-Assam, et al., 2011) where normally the $LH_3$ outperform even the $LL_3$. This seems to indicate that the degradation in the case of CCTV cameras has a more adverse impact of higher frequency coefficients, associated with image features, and in particular, the horizontal features. The SR images of SCface can be seen to have much more artefacts (mostly in the horizontal directions, see Figure 6.12 & Figure 6.13), leading to the poor quality of local features and in particular with the horizontal facial features associated with significant coefficients in $LH_3$ subband being more affected by the ge-

ometrical degradation than vertical features, thus, reducing the accuracy. While the $LL_3$ subband feature gives the best results, thereby being a more suitable face descriptor for recognition under an uncontrolled environment since $LL_3$ is the approximation of the spatial domain image.

## Rank-based Identification

For a given rank $N > 0$ the recognition rate, represent the probability of correctly identifying an individual from a gallery. The above results show improved accuracy rates for recognition of CCTV images when SR-dictionaries are used as compared to existing schemes (see (Grgic, et al., 2011)). In this section, we present the results of investigations to determine accuracy rates for a range of ranks (ranging from rank1 to rank 15) identification when the various SR-dictionaries are used on face images from the SCface database surveillance cameras with varying resolution and poses. The experimental results, shown below in Figure 6.8, demonstrate that the identification recognition rates are increased at each distance and the accuracy rate reaches around 54% from 15% in distance 1 when the $LL_3$ wavelet subbands are used as the feature vector.

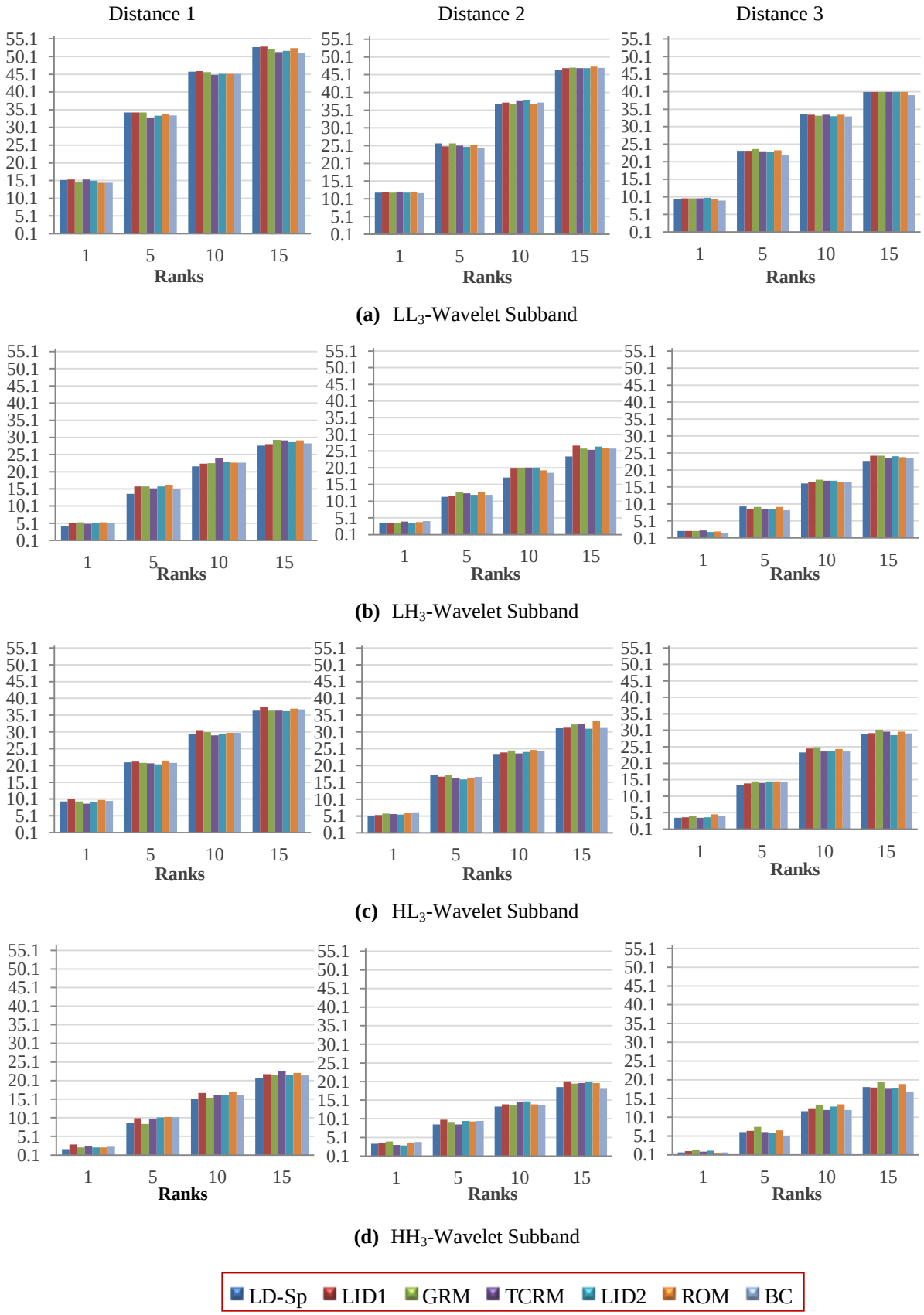



**Figure 6.8:** Rank-N identification accuracy rates for SR images from different resolution enhancement methods

Although recognition with a rank higher than 5 may seem undesirable, however when dealing with recognition in a surveillance scenario often we can benefit from cross checking with recognition from videos captured by different CCTV cameras in nearby locations.

### 6.4.2 Face Recognition Experiments – Configuration 2

For the second dataset configuration, Figure 6.9 and Figure 6.10 show identification accuracy rates when the enrolment reference images are from distance 2 or from distance 3 respectively and the probe image captured is always far from the camera (i.e. at distance 1). We can see from Figure 6.9 that there is very slight difference in terms of recognition between the SR dictionaries and the standard bi-cubic interpolation method. On the other hand the random dictionary GRM in some cases (when $LH_3$ & $HH_3$ are used as feature vectors) can improve accuracy rates more than the image training-based dictionary. The results also demonstrate that CS based SR methods produce better identification accuracy rates when the gallery image set contains images from a distance 2 than distance 3 (i.e. the gallery set contains images from distance 3). In spite of this the improved results in rank one (around 24%) which higher than the performance of the non-dictionary SR method referred to in the introduction (Peng, 2011) where the achieved recognition rates are around 10% to 15%.

Generally speaking, SCface is a very difficult database and from our experiments, it can be seen that the proposed $LID_1$ dictionary based on the SR method can slightly improve recognition performance and performs better than the existing dictionary which is based on training images. In addition the performance of the SR method in terms of face recognition in distance 3 is less than other distances due to the pose variation problem.

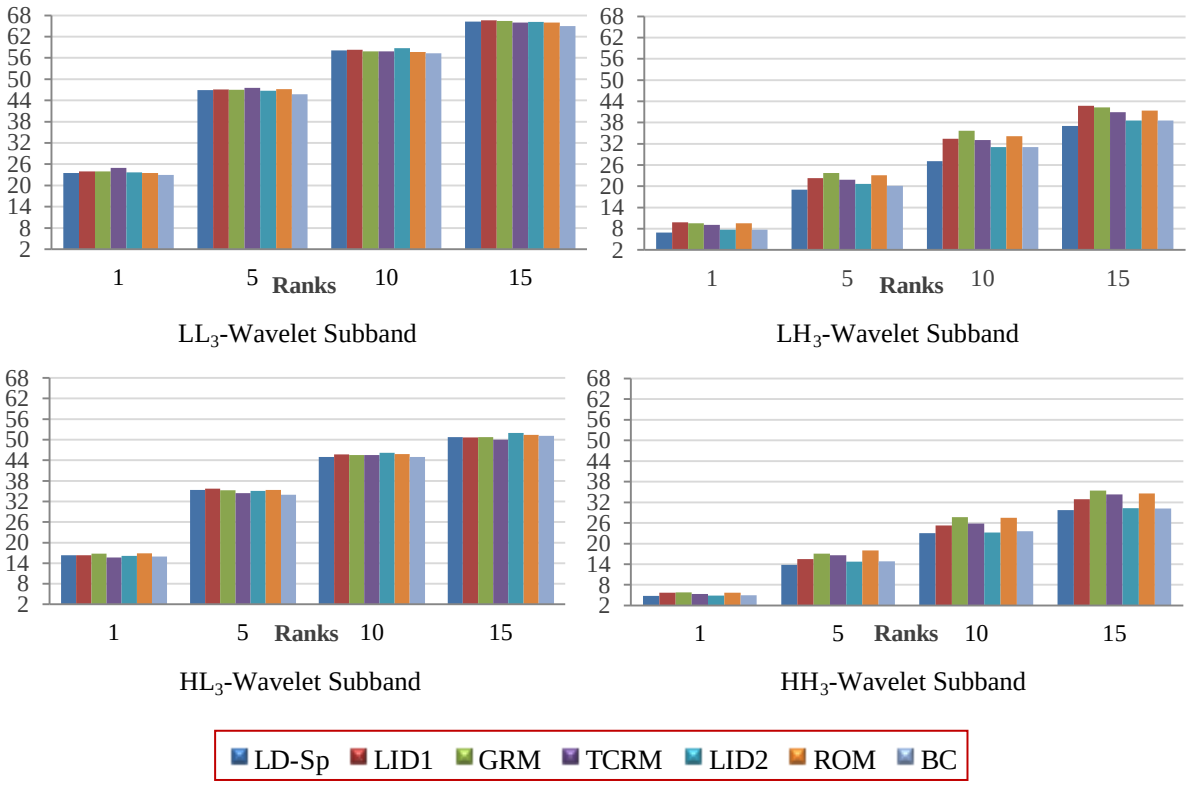



**Figure 6.9:** Rank-N recognition rates for the SCface database, SR dictionaries and bi-cubic method used to enhance LR images from a distance one where the gallery images from distance 2.

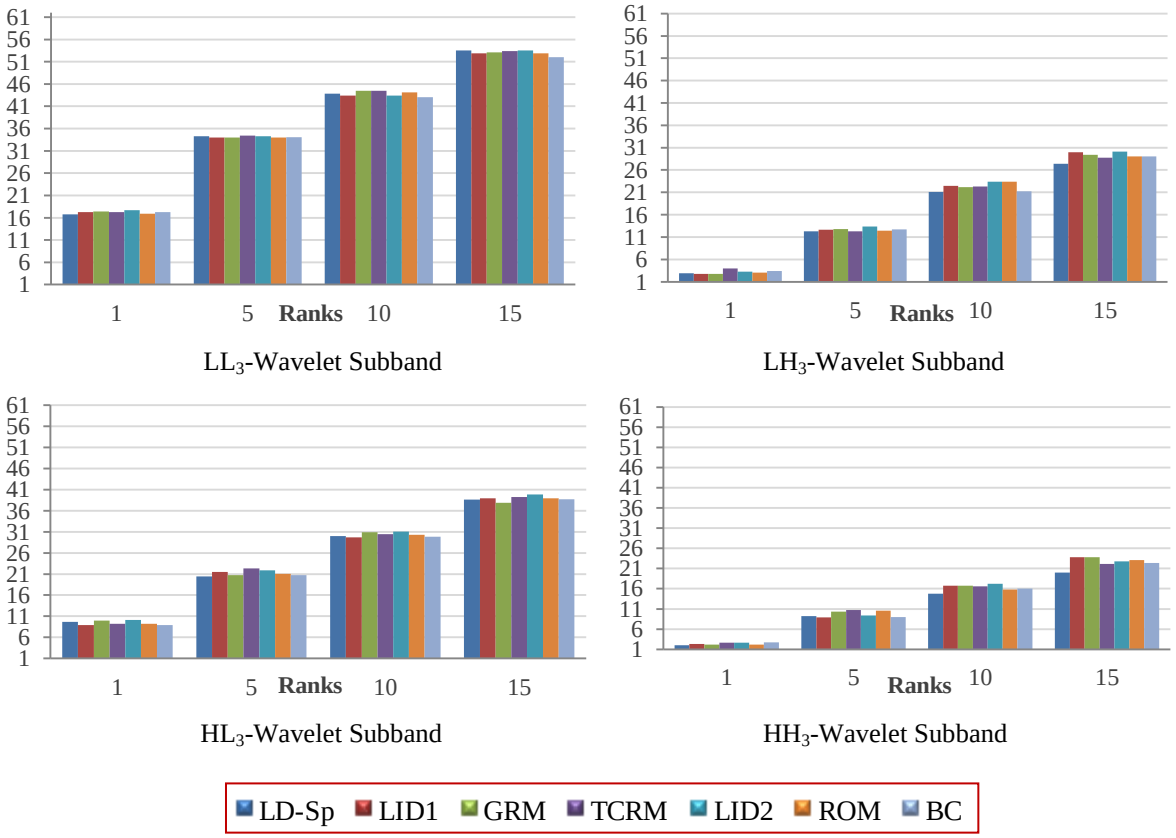



**Figure 6.10:** Rang-N recognition rates for the SCface database, SR dictionaries and bi-cubic method used to enhance LR images from a distance one where the gallery images from distance 3.

## Fusion of Different Subbands

Multi-stream face recognition approaches that fuse different subbands of wavelet transformed face images under varying illumination (Sellahewa & Jassim, 2010), (Abboud, 2011), (Jassim & Sellahewa, 2005) have been shown to improve face recognition. In (Sellahewa & Jassim, 2010) proposed adaptive fusion method for improving face recognition, which demonstrated that the best improvement in recognition accuracy was achieved when LL and LH subband scores are fused. Based on the above results in Figure 6.7 as well as in Figures (6.9 and 6.10) that demonstrated the single stream low frequency approximation subband is a suitable face descriptor for recognition. In this chapter, fusion of match scores (rank one recognition) from approximation and detail subbands for the super resolved image (i.e. after reconstructing an HR image by using CS based SR methods) has been investigated to observe the effect of the multi-stream fusion approach on identification accuracy rate. Here several fusion parameters are selected using fixed weights. Our experiments have been conducted on the second dataset configuration and the identification accuracy rates based on score fusion of $LL_3$ with $LH_3$ subband are given in Table 6.1 and Table 6.2. The experimental results showed that the multi-stream fusion of the two subbands slightly outperform the best single stream ($LL_3$ subband) when the higher weight is given to the low-frequency subband ($W_{LL_3}$=0.9) and the lower weight is given to the horizontal frequency subband ($W_{LH_3}$=0.1), but this improvement is not significant.

| Gallery set | Probe set | $LL_3$+$LH_3$ | | SR methods | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | $W_{LL3}$ | $W_{LH3}$ | LD-Sp | GRM | TCRM | LID1 | ROM | LID2 | BC |
| Distance 2 | Distance 1 | 1 | 0 | 23.53 | 24 | 24.92 | 24 | 23.53 | 23.69 | 23.07 |
| | | **0.9** | **0.1** | **24** | **24.61** | **24.76** | **24.15** | **24.61** | **24.61** | **23.38** |
| | | 0.8 | 0.2 | 23.53 | 23.84 | 23.69 | 24 | 23.69 | 23.38 | 23.38 |
| | | 0.7 | 0.3 | 22.61 | 22.46 | 22.61 | 22.15 | 22.61 | 22.61 | 22.15 |
| | | 0.6 | 0.4 | 20.92 | 22.15 | 22.15 | 22 | 21.69 | 20.46 | 20.92 |
| | | 0.5 | 0.5 | 19.23 | 20.61 | 20.15 | 19.84 | 20.30 | 18.76 | 19.38 |
| | | 0.4 | 0.6 | 16.61 | 18.76 | 18.15 | 18.76 | 18.30 | 17.38 | 17.84 |
| | | 0.3 | 0.7 | 13.84 | 18 | 16 | 17.23 | 17.23 | 15.23 | 14.92 |
| | | 0.2 | 0.8 | 11.53 | 14.46 | 13.07 | 13.38 | 13.38 | 12.15 | 12.76 |
| | | 0.1 | 0.9 | 9.07 | 11.53 | 11.38 | 11.69 | 11.84 | 11.07 | 11.38 |
| | | 0 | 1 | 6.92 | 9.53 | 9.07 | 9.84 | 9.53 | 7.69 | 8.76 |

**Table 6.1:** Rank one identification accuracy (%) for the LL and LH based multi-stream subband fusion approach of super resolution images from LR Probe images of distance 1. Gallery set contains distance 2 images.

| Gallery set | Probe set | $LL_3$+$LH_3$ | | SR methods | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | $W_{LL3}$ | $W_{LH3}$ | LD-Sp | GRM | TCRM | LID1 | ROM | LID2 | BC |
| Distance 3 | Distance 1 | 1 | 0 | 16.76 | 17.38 | 17.23 | 17.23 | 16.92 | 17.69 | 17.23 |
| | | **0.9** | **0.1** | **17.23** | **17.23** | **17.84** | **17.38** | **17.69** | **18** | **17.53** |
| | | 0.8 | 0.2 | 17.07 | 16.61 | 16.76 | 17.07 | 17.07 | 17.23 | 17.38 |
| | | 0.7 | 0.3 | 16.15 | 16 | 16.15 | 16.46 | 16.30 | 16.46 | 16.15 |
| | | 0.6 | 0.4 | 14.30 | 14.15 | 14.15 | 14.15 | 14.46 | 14.46 | 15.07 |
| | | 0.5 | 0.5 | 12.46 | 12.92 | 12.76 | 12 | 12.61 | 12.30 | 12.76 |
| | | 0.4 | 0.6 | 10 | 10 | 10.92 | 10.30 | 10.15 | 10.46 | 10.92 |
| | | 0.3 | 0.7 | 8.15 | 8.46 | 9.07 | 8.15 | 8.30 | 8.61 | 8.76 |
| | | 0.2 | 0.8 | 5.69 | 6.30 | 6.61 | 6.46 | 6.30 | 6.30 | 6.46 |
| | | 0.1 | 0.9 | 4.30 | 4.92 | 5.38 | 3.84 | 4.61 | 4.76 | 4.92 |
| | | 0 | 1 | 2.92 | 2.76 | 4 | 2.76 | 3.07 | 3.23 | 3.38 |

**Table 6.2:** Rank one identification accuracy (%) for the LL and LH based multi-stream subband fusion approach of super resolution images from LR Probe images of distance 1. Gallery set contains distance 3 images**.**

We repeated the experiment by fusing the $LL_3$ subband with the $HL_3$ at the score level, and the results are displayed in 6.3 and 6.4 Tables. As can be seen again the best performance is achieved with the weighting combination ($W_{LL_3}$= 0.9 and $W_{HL_3} = 0.1$), and the results for this combination is marginally different from those obtained when fused the $LL_3$ with the $LH_3$ subbands. However, for all the other weighting combinations for which $W_{HL_3} > 0$, the fusing of $LL_3$ and $HL_3$ outperforms the scheme that fuses $LL_3$ and $LH_3$. This reflects the fact that the single $HL_3$ subband scheme outperforms the $LH_3$ subband scheme possibly because of the observed degradation of the images in the horizontal directions.

| Gallery set | Probe set | $LL_3$+ $HL_3$ | | SR methods | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | $W_{LL3}$ | $W_{HL3}$ | LD-Sp | GRM | TCRM | LID1 | ROM | LID2 | BC |
| Distance 2 | Distance 1 | 1 | 0 | 23.53 | 24 | 24.92 | 24 | 23.53 | 23.69 | 23.07 |
| | | **0.9** | **0.1** | **24.15** | **23.84** | **24.92** | **23.53** | **24.15** | **23.69** | **23.23** |
| | | 0.8 | 0.2 | 23.38 | 23.84 | 24.46 | 23.69 | 23.38 | 23.38 | 22.76 |
| | | 0.7 | 0.3 | 22.76 | 23.38 | 23.23 | 23.69 | 23.07 | 22.92 | 22.76 |
| | | 0.6 | 0.4 | 22 | 22.92 | 22.46 | 22.30 | 22.92 | 22.61 | 22.30 |
| | | 0.5 | 0.5 | 22.61 | 22.76 | 22.15 | 22.30 | 21.84 | 22.46 | 22 |
| | | 0.4 | 0.6 | 21.84 | 21.53 | 22.46 | 21.53 | 21.38 | 21.84 | 22.15 |
| | | 0.3 | 0.7 | 20.76 | 20.61 | 20.92 | 20.30 | 20.76 | 21.38 | 21.07 |
| | | 0.2 | 0.8 | 19.84 | 19.53 | 19.38 | 20 | 19.69 | 20.30 | 20.92 |
| | | 0.1 | 0.9 | 18.15 | 18.61 | 17.84 | 18.30 | 18.15 | 18.61 | 18.46 |
| | | 0 | 1 | 16.30 | 16.76 | 15.69 | 16.30 | 16.92 | 16.15 | 16.30 |

**Table 6.3:** Rank one identification accuracy (%) for the LL and HL based multi-stream subband fusion approach of super resolution images from LR Probe images of distance 1. Gallery set contains distance 2 images.

| Gallery set | Probe set | $LL_3$+ $HL_3$ | | SR methods | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | $W_{LL3}$ | $W_{HL3}$ | LD-Sp | GRM | TCRM | LID1 | ROM | LID2 | BC |
| Distance 3 | Distance 1 | 1 | 0 | 16.76 | 17.38 | 17.23 | 17.23 | 16.92 | 17.69 | 17.23 |
| | | **0.9** | **0.1** | **17.38** | **17.53** | **17.84** | **17.53** | **17.53** | **18** | **17.38** |
| | | 0.8 | 0.2 | 17.38 | 17.38 | 17.07 | 17.38 | 17.23 | 17.69 | 17.53 |
| | | 0.7 | 0.3 | 16.76 | 17.38 | 17.53 | 17.07 | 17.07 | 16.76 | 17.38 |
| | | 0.6 | 0.4 | 16.76 | 16.76 | 16.92 | 16.61 | 16.92 | 16.76 | 16.61 |
| | | 0.5 | 0.5 | 16.46 | 16.61 | 16.61 | 16.76 | 16.92 | 16.61 | 17 |
| | | 0.4 | 0.6 | 15.53 | 16.46 | 15.23 | 15.38 | 15.84 | 15.53 | 16 |
| | | 0.3 | 0.7 | 14.30 | 13.84 | 14 | 14.15 | 14.15 | 14.15 | 14.30 |
| | | 0.2 | 0.8 | 13.07 | 13.38 | 13.23 | 12.92 | 13.23 | 12.92 | 13.53 |
| | | 0.1 | 0.9 | 11.69 | 12.30 | 12 | 12 | 11.69 | 11.84 | 11.84 |
| | | 0 | 1 | 9.69 | 10 | 9.23 | 8.92 | 9.23 | 10.15 | 8.92 |

**Table 6.4:** Rank one identification accuracy (%) for the LL and HL based multi-stream subband fusion approach of super resolution images from LR Probe images of distance 1. Gallery set contains distance 3 images.

Overall, the multi-stream fusion of the two subbands of the super resolution images from surveillance input images illustrates the conclusion that the improvement of the performance of face recognition using the multi-stream approach is slight, if any.

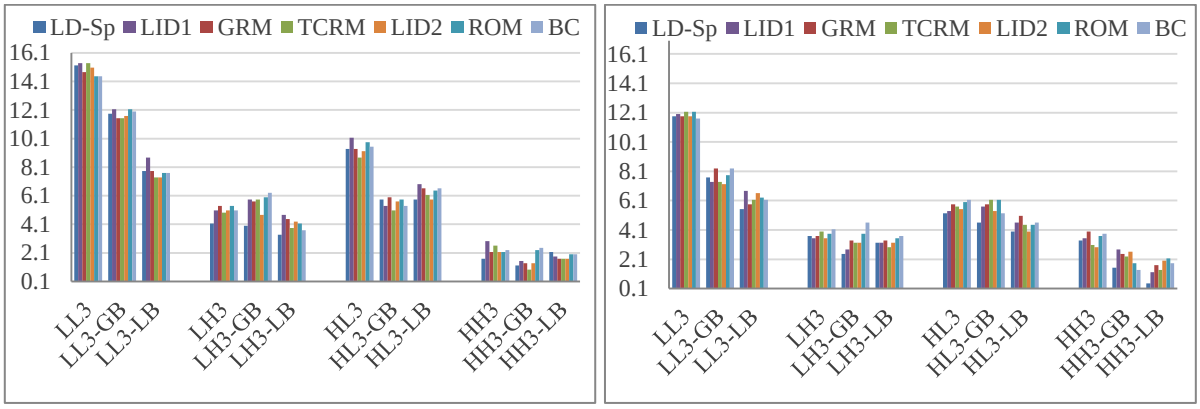
### 6.4.3 Recognition with Binary Feature Vectors

In all the above experiments, the coefficients in each wavelet subbands of SR images after normalisation are used as a feature vector for recognition purposes. However, binaries wavelet coefficients with error correction codes (Hussein, 2014) have been used as a preprocessing technique for identifying individuals, and studied to detect and correct errors in binary feature vectors (which resulted from variation in recording conditions such as illumination conditions) and improve face recognition. In this chapter, different binarization approaches have been applied to observe the effect of the binaries wavelet feature vector of the reconstructed super resolve image for identification accuracy rates.

There are a number of ways of binarising images, but here a global binary approach based on global threshold (i.e. global mean of each subband), and local binary approach that is based on a local threshold to each 3 × 3 overlap blocks (i.e. local mean to each block) of the wavelet coefficient subband are tested. Global and/or Local binaries, single-stream wavelet subbands $I$ are calculated according to the following formula:

$$\boldsymbol{B(i,j)} = \begin{cases} \mathbf{1} & \boldsymbol{if\ I(i,j) \geq \mu} \\ \mathbf{0} & \boldsymbol{Otherwise} \end{cases}$$

Where $\mu$ is the global or local mean feature vector and the hamming distance is used to calculate the distance between pairs of feature vectors corresponding to the same subbands. The binaries face feature vectors (even without using error correction codes), are shown to generally perform better than the non-binaries feature vectors for other kinds of image degradation due to illumination or mild blurring effects (Hussein, 2014). However, the results of our experiments demonstrate that binaries feature vectors do not seem to be suitable for the reconstructed image by different SR approaches. In general, binaries feature vectors do not give better results than the non-binaries feature vectors. Recognition accuracy rates of the different approaches for binary coefficients of the super resolved image are given in Figure 6.11, below. Note that, here we are not applying error correction and it is possible that error correction may result in improved performance. However, this would require an investigation into the error model associated with images from CCTVs in terms of distances and other recording conditions.




(a) Distance 1

(b) Distance 2

(c) Distance 3

**Figure 6.11:** Shows recognition accuracy (%): Global / and Local binaries feature vectors of the reconstructed super resolve images are applied where the Low-resolution-SCface images at varying distances and the gallery set contains high-resolution frontal images.

## 6.5 Post-Super Resolution Image Quality

The second stage of the experiments is to compare the performance of SR dictionaries that do not need training images with the existing LD dictionary in the spatial domain that depend on face image information and with the well-known bi-cubic interpolation method in terms of enhancing the resolution of the input LR image from three different distances. The reconstruction fidelity for our super resolved face images in the SCface database was quantified using four measures quality such as adaptive quality measure that not depends on HR reference image i.e. Histogram intersection as described in Chapter 4, PSNR, Contrast, and Correlation. The HR images from mug-shot camera are used as the ground truth to calculate quality values. Table 6.5 below shows the average values per measure face image quality. It can be seen from Table 6.5 that there is no significant difference between SR by different dictionaries independent of the tested images and bi-cubic method. However, quality values by Histogram intersection showing different results, where at each distance ranges the SR by different dictionary methods are able to improve image quality higher than bi-cubic method. Moreover, the LD-Sp dictionary produces lower quality images at each distance (i.e. close, middle & far distance). In general, our experiments show that the dictionaries, which are not dependent on images slightly, improve image resolution.

| Gallery | Probe images | SR methods | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | | LD-Sp | GRM | TCRM | $LID_1$ | ROM | $LID_2$ | BC |
| Frontal images 128x128 | Distance 1 | 0.69 | 0.71 | 0.71 | 0.71 | 0.71 | 0.71 | 0.70 |
| | Distance 2 | 0.70 | 0.73 | 0.73 | 0.72 | 0.73 | 0.72 | 0.72 |
| | Distance 3 | 0.63 | 0.67 | 0.66 | 0.66 | 0.66 | 0.66 | 0.65 |

(a)

| Gallery | Probe images | SR methods | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | | LD-Sp | GRM | TCRM | $LID_1$ | ROM | $LID_2$ | BC |
| Frontal images 128x128 | Distance 1 | 0.17 | 0.20 | 0.20 | 0.20 | 0.21 | 0.20 | 0.21 |
| | Distance 2 | 0.13 | 0.16 | 0.15 | 0.15 | 0.16 | 0.15 | 0.16 |
| | Distance 3 | 0.12 | 0.15 | 0.14 | 0.14 | 0.15 | 0.14 | 0.15 |

(b)

| Gallery | Probe images | SR methods | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | | LD-Sp | GRM | TCRM | $LID_1$ | ROM | $LID_2$ | BC |
| Frontal images 128x128 | Distance 1 | 12.50 | 12.83 | 12.80 | 12.80 | 12.83 | 12.75 | 12.28 |
| | Distance 2 | 12.41 | 12.87 | 12.83 | 12.81 | 12.83 | 12.76 | 12.27 |
| | Distance 3 | 11.19 | 11.76 | 11.69 | 11.68 | 11.70 | 11.64 | 12 |

(c)

| Gallery | Probe images | SR methods | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | | LD-Sp | GRM | TCRM | $LID_1$ | ROM | $LID_2$ | BC |
| Frontal images 128x128 | Distance 1 | 11.83 | 12.19 | 13.50 | 11.86 | 12.58 | 13.01 | 7.21 |
| | Distance 2 | 11.42 | 12.01 | 12.15 | 11.42 | 12.18 | 12.26 | 9.27 |
| | Distance 3 | 8.83 | 9.98 | 9.30 | 9.59 | 10.47 | 10.58 | 10.14 |

(d)

**Table 6.5:** Quality values of SR images: (a) Contrast measure, (b) Correlation measure, (c) PSNR measure, and (d) Histogram Intersection measure.

Finally, for more testing about the viability of our overcomplete SR dictionaries that constructed without using face images, but satisfy CS properties, to overcome the problems of images captured at distances from lower quality/resolution web cameras in terms of super resolve image quality. We selected randomly two different subjects from the SCface image database as presented below in Figures 6.12 and 6.13, where the face images captured at different distances, close (distance 3), mid (distance 2), far (distance 1) from surveillance camera 1. As can be seen the difference between the recovered HR images by using SR dictionaries independent on images is slightly noticeable by the human eye and a significant improvement in image quality can be seen when the $LID_1$ dictionary is used. Although there is a small difference between the dictionary methods in terms of quality, TCRM produced more artifacts than GRM and reduced slightly the quality of the image. On the other hand, despite of the SR method by LD-Sp dictionary that depends on the training image set being able to recover the HR image, but this method generates more aliasing around the mouth and eyes, and that affect on the image quality. Therefore, the super resolved image by $LID_1$ is better than the one enhanced by the existing dictionary LD-Sp although the $LID_1$ as noticeably generates a relatively blurred image. In addition, the quality of the $LID_1$ images is slightly higher than bi-cubic interpolation images.

In summary, the above results reveal a significant improvement noticeable in SR images over the LR images as well as a greater improvement in resolution from the SR by $LID_1$ dictionary.

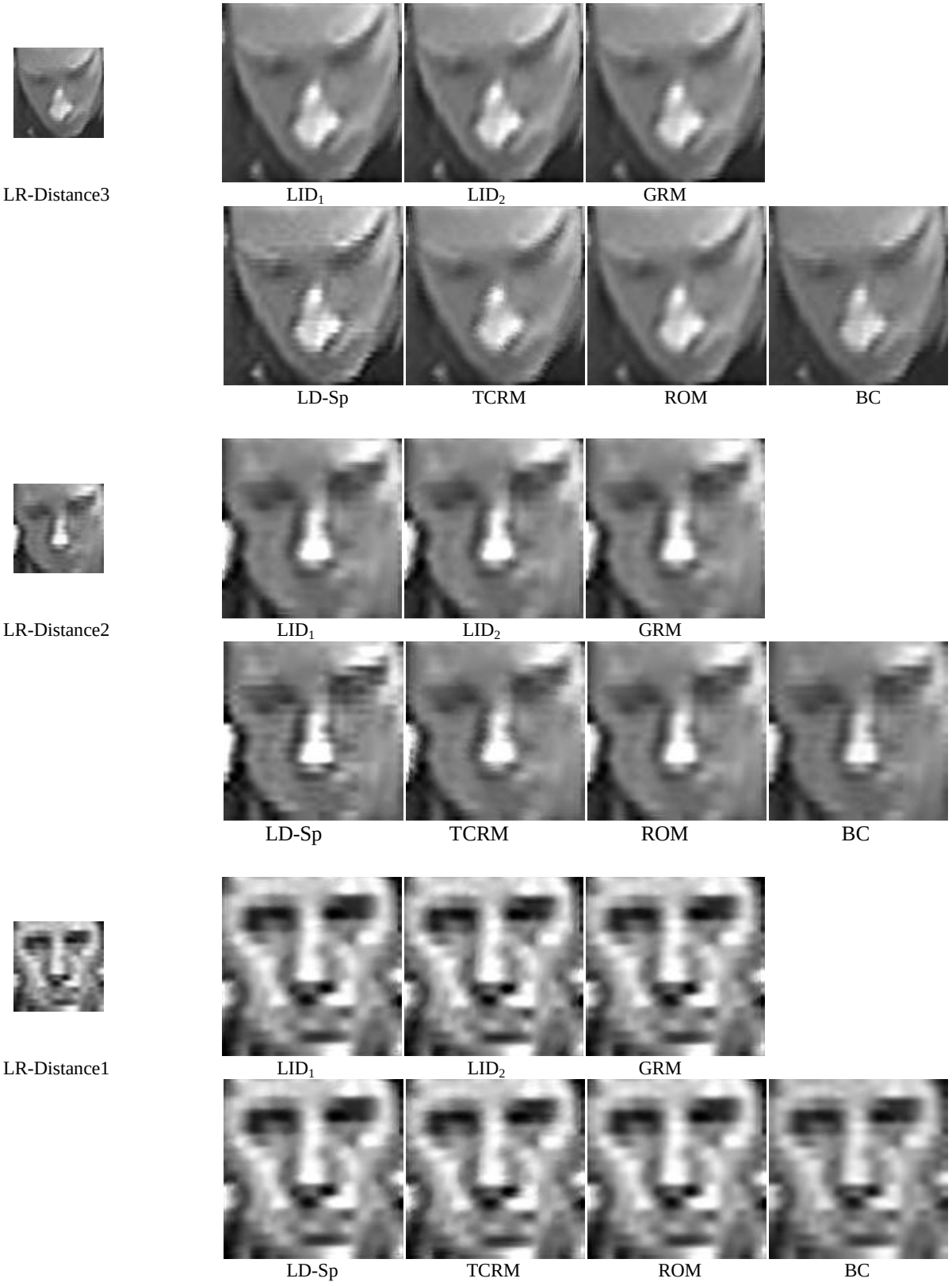



**Figure 6.12:** Example 1, Comparison between SR approaches by different dictionaries and bi-cubic method to reconstruct the low-resolution images at three different distances from SCface database.

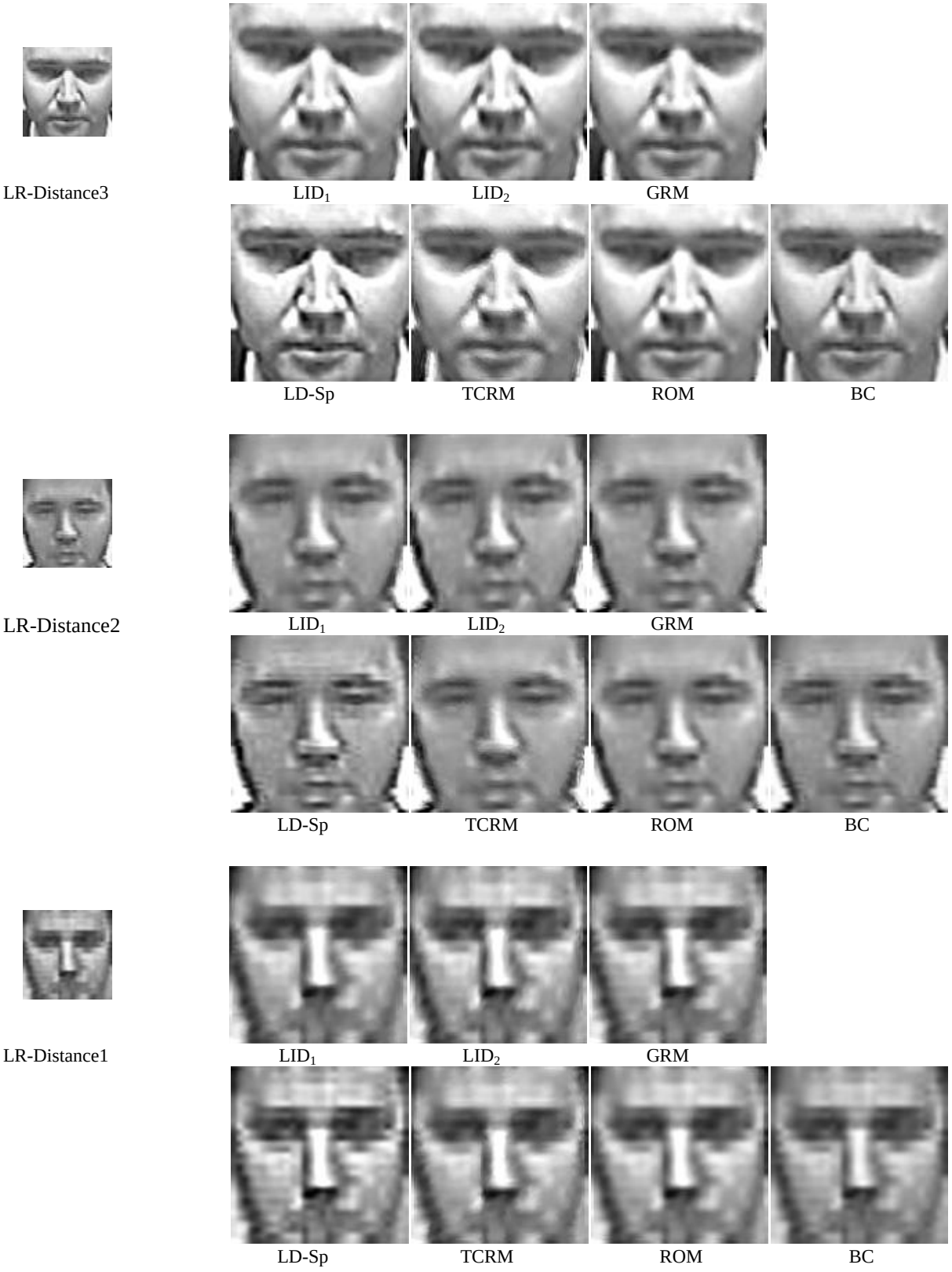



**Figure 6.13:** Example 2, comparison between SR approaches by different dictionaries and bi-cubic method to reconstruct the low-resolution images at three different distances from SCface database.

## 6.6 Summary and Conclusion

Face images captured from surveillance video are usually of low quality and therefore it is hard to provide detailed information of face features. Super resolution based CS dictionary approaches have been presented to reconstruct an HR image from a single image for improving the recognition performance of LR images obtained from CCTV cameras. Comprehensive experiments on images from SCface database that mimic real world conditions are performed. Our results show that the SR dictionaries which do not need training images are suitable for application to surveillance footage and on the one hand are superior when compared to matching in LR process (Grgic, et al., 2011) whilst on the other hand are slightly higher than learning dictionaries in some cases. The artefacts that generated as a result, visually distracting to humans and are more effective on machine recognition algorithms. The rank-one identification recognition rates are still likely to be poor, despite the improvement provided by CS dictionary methods.

Image super resolution should be carried out by using some a priori information about the degradations in the input LR images where with this information the solution to the reconstruction problem can carry out easily. However, this information in general is not available for the surveillance image at a distance. Therefore, there is a need to develop a model of degradation that encapsulate the distortion to the frequency domain of images captured by CCTV cameras at different distances, instead of using the approximation approach described above. The estimation of frequency domain distortion can be used for enhancements (i.e. remove distortion/blurring from the observed image) through inverting the appropriated frequency domain filter instead of the restoration technique that uses Weiner filter.

# Chapter 7

## CONCLUSIONS AND FUTURE RESEARCH

### 7.1 Conclusions

Face recognition is one of the most desirable biometric-based identification or verification systems. It has a number of real world applications such as surveillance systems, law enforcement and airport security, where it is an important tool in crime fighting and robbery. The performance of face recognition degrades considerably when the input images are of low-resolution, of small size and have degraded quality, as is often the case for images taken by surveillance cameras or from a distance, where the decrease in resolution leads to losing discriminatory properties that distinguish one person from another.

In this thesis, we investigated the problems related to degrade low-resolution face images of limited quality and how their influences on the whole face recognition system can be addressed by super resolution techniques based on the recently developed compressive sensing theory. The main aims of this thesis are to study and understand the properties of compressive sensing and use the gained knowledge to (1) develop various approaches to designing new super resolution dictionaries that satisfy the compressive sensing conditions and can be used to overcome the resolution limitation in the probe input images and create images with high-resolution. (2) Test the performance of different overcomplete dictionaries such as exiting matrices and the proposed dictionaries to address the consequences of face recognition under uncontrolled conditions, and (3) test the need for using image sets which rely on a sufficiently large and representative range of face images for a sample of persons for constructing learning dictionaries. The work done and the achievements of this thesis can be summarized as follows:

1. At the outset of this study, we reviewed the existing work in the literature regarding super resolution techniques; compressive sensing and the sparse recovery based super resolution methods. The non-CS based iterative optimization method and the overcomplete

dictionaries constructed from image patches belonging to certain face image dataset have studied to improve image quality in the spatial domain from single or multiple degraded low-resolution images with different strengths of blur degradation. Moreover, we have also investigated the impact of super resolution methods to improve image resolution in the frequency domain by enhancing wavelet coefficients of the input low-resolution image. The experiments demonstrated that the super resolution techniques in the spatial domain achieve better results in terms of measured quality. These results raised the following question: are the existing random dictionaries which are constructed without using image database and guarantee the recovery of unique $k$-sparse signal able to improve image super resolution; this was investigated next.

2. We proposed overcomplete super resolution dictionaries constructed based on compressive sensing theory and not depend on face image information as well as the various known random matrices investigated to recover image with good quality from the degraded low-resolution image, and the images patches dictionaries have been used for comparison. We first demonstrated that, the different super resolution dictionary methods in terms of image quality are able to reconstruct high-resolution images and produce superior results than the iterative method and bi-cubic resolution enhancement technique. Secondly, the performance of various dictionaries decreases with an increased level of blurring of the input low-resolution image. Thirdly, the random and proposed dictionaries produce similar, if not more, improvement than the existing dictionaries that depend on image patches, which leads us to the most important conclusion of removing the need face image information in the training sets to construct dictionaries in terms of super resolution image quality.

3. Motivations by the above observations and conclusions, we have investigated the use various super resolution dictionaries to construct full-face images in the spatial domain for face recognition under uncontrolled conditions and at a distance. The results demonstrated that the various types of dictionaries could improve the performance of face identification and the pattern of accuracy is not depend on the feature extraction schemes, which the accuracy reached 100% for the LH, HL, and HH wavelet subbands in subset 2 and about 100% for PCA for one subset. In addition, the accuracy rates are not affected by different geometrical blurring degradation in the low-resolution image where the identification accuracy rates remain at the same level. However, the standard

interpolation technique could achieve best accuracy rates than the SR methods when the input low-resolution with slightly good quality (i.e. with low degradation). Therefore, the success of SR dictionary methods can be apparent when the low-resolution images are of poor quality. Moreover, the performance of dictionaries that do not rely on image patches firstly can yield significant improvement in accuracy rates than the matching in low-resolution without using super resolution techniques to the probe images and are much more apparent as the image quality deteriorates from mild to severe degradation. Secondly, it produces similar if not slightly better than the learnt images dictionaries. This confirms that the training images set to construct dictionaries are not necessary for face recognition systems.

4. We have proposed learning overcomplete dictionaries that are generated based on the wavelet coefficients subbands of the training face images set as well as the various overcomplete dictionaries that are constructed in the spatial domain have investigated to super resolve the feature vectors that are extracted from the degraded low-resolution image for face recognition. Two models of blurring with different parameters used to obtain the degradation of the low-resolution image. We first demonstrated that the performance of the proposed learning dictionaries in the wavelet domain outperform than the various overcomplete dictionaries in the spatial domain including the non-adaptive dictionaries that constructed without image information and the non-adaptive dictionaries are not suitable for use in the frequency domain. However, the significant improvement in terms of identification accuracy rates can be achieved by using different compressive sensing dictionary methods for super resolve full-face image in the spatial domain. Where the accuracy rate for the various subbands in subset two ranges from a minimum of 78% to a maximum of 100% while, this range dropping to 46% (minimum) – 87% (maximum) when the different dictionaries are used to recover the feature vector of the LR images with severe degradation. Secondly, the level of blur degradation in the input LR image has a significant effect on the performance of SR dictionary methods to reconstruct the feature vectors $LH_3$ and $HL_3$ where the identification accuracy rates decrease with increased levels of blurring. Thirdly, there is a little difference in identification accuracy rate between the various SR dictionaries in the spatial domain.

   On the other hand, SR dictionaries in both domains have also investigated to reconstruct the feature vector of the LR images from standard definition video camera for face recognition. Our results have demonstrated that the CS based super resolution dictionar-

ies in the wavelet domain could obviate the need for costly high definition cameras for face recognition at a distance by super resolved $LH_3$ wavelet subband of the images captured by less expensive standard definition cameras.

5. Necessary and/or sufficient structural compressive sensing conditions of the overcomplete matrices that are relevant to the recovery of a sparse image from the degraded low-resolution image have been studied. The Restricted Isometry Property has been investigated for various overcomplete dictionaries and a number of statistical tests of numerical matrix parameters relevant to this condition have been conducted/and demonstrated that. The existing dictionary that is learned from face image is a highly ill-conditioned matrix where the condition number is high, equalling 1.00E+15 and far from satisfying the property of compressive sensing where the row rank condition is low, therefore the probability of finding linearly dependent columns is high in this matrix. However, the new dictionaries that do not depend on face image information are well-conditioned matrices and satisfy the necessary and sufficient compressive sensing property, where they have a reasonably low-condition number between $[1.48, 1.84]$ and with full row rank.

6. The earlier results inspired us to go further to investigate the influence of compressive sensing and super resolution techniques for addressing surveillance images from CCTV camera's, where the quality of the surveillance images are affected by a complex set of factors including variations in resolutions, illumination and poses. The distortion/degradation of CCTV images in general is unavailable and difficult to model as well as these images have poor contrast. Hence, in this thesis contrast measure and environmental degradation function was investigated to approximate the unknown degradation model for super resolution schema. Our results confirmed and provided strong evidence that the compressive sensing theory can be able to develop a face recognition scheme without using a training image set to construct super resolution dictionaries. The overcomplete dictionaries that do not need training images produced better results in terms of quality measure and in identification accuracy rates than learnt dictionaries based on a training database and also than the standard interpolation method. In addition, the identification accuracy reached 65% in rank15 when the LL wavelet subband was used as the feature vector, which is higher than the existing matching without using the SR method.

## 7.2 Future Research Directions

The work reported in this thesis, not only demonstrated the viability of non-adaptive compressive sensing dictionaries & super resolution to overcome the problem of small size low-resolution images and improve face recognition under uncontrolled conditions. However, also highlights several potential research directions to be explored in the future. Some examples of future research directions are listed below:

- More investigations and experiments need to be conducted to reconstruct a high-resolution image from a single or multiple low-resolution images by using another class of structured random dictionaries that does not depend on the training images such as Random Partial Fourier matrices and Bernoulli Random matrices for face recognition.

- Uncontrolled condition may arise due to facial expression changes from one or more motion of the muscles of the face such as happiness, fear, anger, sadness or surprise. Where humans are capable of showing their feelings and may be unaware that the surveillance cameras are recording their facial images. These motions adversely affect face image appearance, which leads to large differences between face images of the same person and deterioration image quality. We shall, in our future work, use super resolution dictionaries to overcome the problems of the different situations of a human face due to the expression variation in uncontrolled conditions.

- The initial results of super resolution technique based on dictionaries that are not dependent on a training image set to overcome the problem of low-resolution image from CCTV cameras for face recognition show promising results. However, identifying an individual from CCTV cameras remains an extremely challenging problem. Therefore, further investigation and experiment is needed to improve identification results. Using super resolution dictionary methods to reconstruct the face regions such as eye, nose & mouth region instead of full image and fusion at the feature level of the feature vectors is extracted from the reconstructed regions to improve face recognition. In this respect, we need to have an open mind in using different feature extraction techniques for different facial regions, and fusion may test at different levels such as score level and decision level.

- The rather disappointing accuracy results achieved when the non-adaptive dictionaries were used to super resolve the wavelet face feature vectors, as compared to the LD-WD, which was trained on wavelet subbands of HR images, reveal the need to develop non-adaptive dictionaries that are compatible with the structure of wavelet subbands.

- The limited set of experiments to make observations on the structure of the Fourier spectrum show the need to develop a model of degradation that encapsulate the distortion to the frequency domain of images captured by CCTV cameras at different distances, instead of using the approximation approach described in chapter 6. The estimation of frequency domain distortion can be used for enhancements (i.e. remove distortion/blurring from the observed image) through inverting the appropriated frequency domain filter instead of the restoration technique that uses Weiner filter. In the future work, we will try to develop models of degradation in the frequency domain based on the degradation of the Fourier spectrum of available LR images by designing/ or developing special filters in the frequency domain for image enhancement.

- Address the problem of face recognition from CCTV cameras by developing the performance of CS based SR method to recover an HR image from more than one LR image of the same subject from different cameras by using image registration technique. Image restoration will be used for fusion or combine two face images, hence the beginning step of CS dictionaries based SR method is to determine the relative shifts between the LR images and each pixel of each LR image is put into one particular image based on the restoration information.

- The problems that image inpainting deals with have many common characteristics with the problem of low resolved and low quality images. An interesting future investigation we intend to conduct would include the development and testing of a hybrid scheme that combines current inpainting solution with CS-based super-resolution techniques. Some existing inpainting techniques (Bertalmio, et al., 2000), (Tschumperle & Deriche, 2005) that used to fill in missing certain region/hole in an image introduced some blurring effects when the missing regions are filled in is large, and since the quality of the inpainted image has a critical impact on the quality at the

final resolution. The author of this thesis strongly believes that the super resolution dictionary methods, which differ from the inpainting techniques where the missing region in the super resolution is spread across all images, could be used to improve image resolution reconstructed from inpainting technique.

- Use compressive sensing dictionaries and super resolution in the spatial domain for other biometric traits. For example, one can investigate the use of compressive sensing based super-resolution for deteriorated fingerprint images in forensic applications. We can also attempt to reconstruct full iris images that are recorded at a distance with low-resolution as well as trying to apply various dictionaries to super-resolve the feature vectors for iris recognition.

# REFERENCES

Abboud, A., (2011). Quality Aware Adaptive Biometric Systems. *PhD Thesis, Buckingham University.*

Al-Assam, H., (2013). Entropy Evaluation and Security Measures for Reliable Single/ Multi-Factor Bioetric Authentication and Biometric Keys. *PhD Thesis, Buckingham University.*

Al-Assam, H., Abboud, A. and Jassim, S., (2011). Exploiting Samples Quality in Evaluating and Improving Performance of Biometric Systems. *International Journal for Digital Society,* 2(2).

Al-Assam, H., Abboud, A., Sellahewa, H. and Jassim, S., (2012). Exploiting Relative Entropy and Quality Analysis in Cumulative Partial Biometric Fusion. *Transactions on Data Hiding and Multimedia Security VIII. Springer-Verlag Berlin Heidelberg,* pp. 1-18.

Al-Azzeh, M., Eleyan, A. and Demirel, H., (2008). PCA-based face recognition from video using super-resolution. *IEEE,* pp. 1-4.

AL-Jawad, N., (2009). Exploiting Statistical Properties of Wavelet Coefficients for Image/Video Processing and Analysis Tasks. *PhD Thesis, Buckingham University.*.

Al-Obaydy, W. and Sellahewa, H., (2011). On using high-definition body worn cameras for face recognition from a distance. *Biometrics and ID Management, Springer-Verlag,* pp. 193-204.

Alves, J. C. and Hussein, M. S., (2012). The Moore-Penrose Pseudoinverse: A Tutorial Review of the Theory. *Brazilian Journal of Physics,* 42(1-2), pp. 146-165.

Bailly-Bailliére, E., Bengio, S., Bimbot, F. and Hamouz, M., (2003). The BANCA database and evaluation protocol. *Audio-and Video-Based Biometric Person Authentication,* pp. 625-638.

Bajwa, W. U., Haupt, J. D., Raz, G. M. and Wright, S. J., (2007). Toeplitz-Structured Compressed Sensing Matrices. *Statistical Signal Processing (SSP), IEEE 14th workshop on,* pp. 294-298.

Bannore, V., (2009). Iterative-Interpolation Super-Resolution Image Reconstruction. *Springer-Verlag Berlin Heidelberg,* pp. 19-50.

## References


Baraniuk, R., Davenport, M., DeVore, R. and Wakin, M., (2008). A simple proof of the restricted isometry property for random matrices. *Constructive Approximation*, 28(3), pp. 253-263.

Bertalmio, M., Bertozzi, A. L. and Sapiro, G., (2001). Navier-stokes, fluid dynamics, and image and video inpainting. *Computer Vision and Pattern Recognition,IEEE Computer Society Conference on*, Volume 1, pp. I-355.

Bertalmio, M., Sapiro, G., Caselles, V. and Ballester, C., (2000). Image inpainting. *Proceedings of the 27th annual conference on Computer graphics and interactive techniques. ACM Press/Addison-Wesley Publishing Co.*, pp. 417-424.

Bilgazyev, E., Efraty, B., Shah, S. and Kakadiaris, I., (2011). Sparse Representation-based Super-Resolution for Face Recognition At a Distance. *Proceedings of the British Machine Vision Conference*, pp. 52.1-52.11.

Biswas, S., Aggarwal, G. and Flynn, P. J., (2011). Pose-robust recognition of low-resolution face images. *Computer Vision and Pattern Recognition (CVPR), IEEE Conference on*, pp. 601-608.

Bovik, A. C., (2009). The essential guide to image processing. *Academic Press, Elsevier.*

Candes, E. J. and Donoho, D. L., (2000). Curvelets: A surprisingly effective nonadaptive representation for objects with edges. *Stanford Univ Ca Dept of Statistics*, pp. 1-10.

Candes, E. and Tao, T., (2006). Near-Optimal Signal Recovery From Random Projections: Universal Encoding Strategies?. *Information Theory, IEEE Transactions on*, 52(12), pp. 5406-5425.

Casazza, P. G., (2013). Finite frames: Theory and applications. *Springer New York.*

Chang, H., Yeung, D. and Xiong, Y., (2004). Super-resolution through neighbor embedding. *Computer Vision and Pattern Recognition, Proceedings of IEEE Computer Society Conference on*, Volume 1, p. 275.

Chaudhuri, S., (2001). Super-resolution imaging. *Springer-Netherlands*, Volume 632.

Chen, S. and Donoho, D., (1994). Basis pursuit. *IEEE*, Volume 1, pp. 41-44.

Chen, Z. and Dongarra, J., (2005). Condition numbers of Gaussian random matrices. *SIAM Journal on Matrix Analysis and Applications*, 27(3), pp. 603-620.

Choi, H. C., Park, U. and Jain, A., (2010). PTZ camera assisted face acquisition, tracking and recognition. *Biometrics: Theory Applications and Systems (BTAS), 4th IEEE,* pp. 1-6.

Chu, J. et al., (2009). Gradient-based adaptive interpolation in super-resolution image restoration. *Signal Processing, ICSP, 9th International Conference on,* pp. 1027-1030.

Daubechies, I., (1990). The wavelet transform, time-frequency localization and signal analysis. *Information Theory, IEEE Transactions on,* 36(5), pp. 961-1005.

Daubechies, I., Grossmann, A. and Meyer, Y., (1986). Painless nonorthogonal expansions. *Journal of Mathematical Physics,* Volume 27, p. 1271.

Donoho, D., (2006). Compressed sensing. *Information Theory, IEEE Transactions on,* 52(4), pp. 1289-1306.

Donoho, D. L. and Elad, M., (2003). Optimally sparse representation in general (nonorthogonal) dictionaries via l1 minimization. *Proceedings of the National Academy of Sciences,* 100(5), pp. 2197-2202.

Duarte, M. and Eldar, Y., (2011). Structured Compressed Sensing: From Theory to Applications. *Signal Processing, IEEE Transactions on,* 59(9), pp. 4053-4085.

Eldar, Y. C. and Kutyniok, G., (2012). Compressed Sensing: Theory and Applications. *Cambridge University Press Cambridge,* Volume 95.

Fanaswala, M., (2009). Regularized Super-Resolution of Multi-View Images. *MSc Thesis, Carleton University.*

Farsiu, S., Robinson, M., Elad, M. and Milanfar, P., (2004). Fast and robust multiframe super resolution. *Image processing, IEEE Transactions on,* 13(10), pp. 1327-1344.

Fei, C., Mei-qing, W. and Choi-Hong, L., (2008). An algorithm for total variation inpainting based on nonlinear multi-grid methods. *Journal of Algorithms & Computational Technology,* 2(1), pp. 15-33.

Fornasier, M., (2010). Numerical methods for sparse recovery. *Theoretical Foundations and Numerical Methods for Sparse Recovery,* Volume 9, p. 93.

Fornasier, M. and Rauhut, H., (2010). Compressive sensing. *Handbook of Mathematical Methods in Imaging. Springer.*

Freeman, W., Jones, T. and Pasztor, E., (2002). Example-based super-resolution. *IEEE Computer Graphics and Applications,* pp. 56-65.

Gabor, D., (1946). Theory of communication. Part 1: The analysis of information. *Electrical Engineers-Part III: Radio and Communication Engineering, Journal of the Institution of,* 93(26), pp. 429-441.

Ganesh, A., Sastry, S. and Ma, Y., (2009). Robust Face Recognition via Sparse Representation. *IEEE Transaction on Pattern Analysis and Maching Intelligence,* 31(2), p. 1.

Gan, L., Ling, C., Do, T. and Tran, T., (2009). Analysis of the statistical restricted isometry property for deterministic sensing matrices using Stein's method. *Preprint.*

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

Gonzalez, C. and Woods, E., (2002). Digital Image Processing. *Pearson Education International. Prentice Hall.*

Grgic, M., Delac, K. and Grgic, S., (2011). SCface-surveillance cameras face database. *Multimedia tools and applications,* 51(3), pp. 863-879.

He, J. and Zhang, D., (2010). Face super-resolution reconstruction and recognition from low-resolution image sequences. *Computer Engineering and Technology (ICCET), 2nd International conference on* , Volume 2, p. 620.

Hennings-Yeomans, P., Baker, S. and Kumar, B., (2008). Simultaneous super-resolution and feature extraction for recognition of low-resolution faces. *IEEE Computer Society Conference on Computer Vision and Pattern Recognition,* pp. 1-8.

Hu, S., Maschal, R. and Young, S. S., (2012). Face recognition performance with superresolution. *Applied Optics,* 51(18), pp. 4250-4259.

Hussein, W., (2014). Error Correction Codes for Face Recognition in Uncontrolled Environments. *PhD Thesis, Buckingham University.*

Irani, M. and Peleg, S., (1991). Improving resolution by image registration. *CVGIP: Graphical Models and Image Processing,* 53(3), pp. 231-239.

Irani, M. and Peleg, S., (1993). Motion analysis for image enhancement: Resolution, occlusion, and transparency. *Journal of Visual Communication and Image Representation,* 4(4), pp. 1993-12.

Jassim, S. A., (2010). Wavelet-Based Face Recognition Schemes. *Face Recognition,Chapter 7, edited by Milos Oravic, INTECH,* pp. 978-953.

Jassim, S. and Sellahewa, H., (2005). A wavelet-based approach to face verification/recognition. *Proceedings of SPIE, the International Society for Optical Engineering*, pp. 598609-1.

Jerri, A., (1977). The Shannon sampling theorem-Its various extensions and applications: A tutorial review. *Proceedings of the IEEE,* 65(11), pp. 1565-1596.

Jialun, K. and Xianghong, T., (2013). An image inpainting method by structural constraints and sample sparse representation. *Signal Processing, Communication and Computing (ICSPCC), IEEE international conference on,* pp. 1-4.

Keys, R., (1981). Cubic convolution interpolation for digital image processing. *Acoustics, Speech and Signal Processing, IEEE Transactions on,* 29(6), pp. 1153-1160.

Kulkarni, N., Nagesh, P., Gowda, R. and Li, B., (2011). Understanding Compressive Sensing and Sparse Representation Based Super Resolution. *Circuits and Systems for Video Technology, IEEE Transactions on,* PP(99), p. 1.

Kutyniok, G., (2013). Theory and Applications of Compressed Sensing. *GAMM-Mitteilungen,* 36(1), pp. 79-101.

Lee, H., Battle, A., Raina, R. and Ng, A., (2007). Efficient sparse coding algorithms. *Advances in neural information processing systems,* Volume 19, p. 801.

Lee, K., Ho, J. and Kriegman, D., (2005). Acquiring linear subspaces for face recognition under variable lighting. *Pattern Analysis and Machine Intelligence, IEEE Transactions on,* 27(5), pp. 684-698.

Levy, D., 2010. Introduction to numerical analysis. *Department of Mathematics and Center for Scientific Computation and Mathematical Modeling, CSCAMM, University of Maryland.*

Li, S., Gao, F., Ge, G. and Zhang, S., (2012). Deterministic Construction of Compressed Sensing Matrices via Algebraic Curves. *Information Theory, IEEE Transactions on,* 58(8), pp. 5035-5041.

Luong, H., Ledda, A. and Philips, W., (2006). Non-local image interpolation. *Image Processing, IEEE International Conference on,* pp. 693-696.

Makwana, R. R. and Mehta, N. D., (2013). Single Image Super-Resolution VIA Iterative Back Projection Based Canny Edge Detection and a Gabor Filter Prior. *International Journal of Soft Computing and Engineering (IJSCE),* 3(1), pp. 2231-2307.

Maturana, M. D. and Soto, A., (2009). Face Recognition with Local Binary Patterns, Spatial Pyramid Histograms and Naive Bayes Nearest Neighbor Classification. *Chilean Computer Science Society (SCCC), International Conference,* pp. 125-132.

Meur, O. L. and Guillemot, C., (2012). Super-resolution-based inpainting. *Computer Vision-ECCV 2012,* pp. 554-567.

Nguyen, K., Fookes, C., Sridharan, S. and Denman, S., (2011). Feature-domain super-resolution for iris recognition. *Image Processing (ICIP), 18th IEEE International Conference on,* pp. 3197-3200.

Parker, J., Kenyon, R. and Troxel, D., (1983). Comparison of interpolating methods for image resampling. *Medical Imaging, IEEE Transactions on,* 2(1), pp. 31-39.

Park, U., (2009). Face Recognition: face in video, age invariance, and facial marks. *PhD Thesis, Michigan State University,.*

Peleg, S., Keren, D. and Schweitzer, L., (1987). Improving image resolution using subpixel motion. *Pattern Recognition Letters,* 5(3), pp. 223-226.

Peng, Y., (2011). Face Recognition at a Distance: a study of super resolution. *MSc. Thesis, University of Twente.*

Phillips, J., Flynn, P., Beveridge, R. and Scruggs, T., (2009). Overview of the multiple biometrics grand challenge. *Advances in Biometrics,* pp. 705-714.

Prince, S. J., Warrell, J., Elder, J. and Felisberti, F. M., (2008). Tied factor analysis for face recognition across large pose differences. *Pattern Analysis and Machine Intelligence, IEEE Transactions on,* 30(6), pp. 970-984.

Rauhut, H., (2010). Compressive sensing and structured random matrices. *Theoretical foundations and numerical methods for sparse recovery,* Volume 9, pp. 1-92.

Rubinstein, R., Bruckstein, A. and Elad, M., (2010). Dictionaries for sparse representation modeling. *Proceedings of the IEEE,* 98(6), pp. 1045-1057.

Sabuncu, M. R., (2004). Entropy-based image registration. *PhD Thesis, Princeton University.*

Sellahewa, H., (2006). Wavelet-Based Automatic Face Recognition for Constrained Devices. *PhD Thesis, Buckingham University*.

Sellahewa, H. and Jassim, S., (2008). Illumination and expression invariant face recognition: Toward sample quality-based adaptive fusion. *Biometrics: Theory, Applications and Systems, BTAS, 2nd IEEE International Conference on*, pp. 1-6.

Sellahewa, H. and Jassim, S., (2009). Image quality-based adaptive illumination normalisation for face recognition. Volume 7306, p. 73061V.

Sellahewa, H. and Jassim, S., (2010). Image-Quality-Based Adaptive Face Recognition. *Instrumentation and Measurement, IEEE Transactions on*, 59(4), pp. 805-813.

Shen, B., Hu, W., Zhang, Y. and Zhang, Y.J., (2009). Image inpainting via sparse representation. *Acoustics, Speech and Signal Processing. ICASSP. IEEE International Conference on*, pp. 697-700.

Shiau, J., Pu, C. and Leu, J., (2010). Pose estimation and conversion to front viewing facial image using 3D head model. *Security Technology (ICCST), IEEE International Carnahan Conference on*, pp. 179-184.

Sochen, N. and Fishelov, D., (2006). Image inpainting via fluid equations. *Information Technology: Research and Education.International Conference on*, pp. 23-25.

Sroubek, F., Kamenicky, J. and Milanfar, P., (2011). Superfast superresolution. *Image Processing (ICIP), 18th IEEE International Conference*, pp. 1153-1156.

Studer, C. S., (2010). Dictionary Learning for Super-Resolution. *Semester Project in Information Technology and Electrical Engineering.*

Tistarelli, M., Li, S. Z., Chellappa, R. and Chellappa, R., (2009). Handbook of Remote Biometrics. *Springer-Verlag London.*

Trefethen, L. N. and Bau III, D., (1997). Numerical linear algebra. *Society for Indstrial and Applied Mathematics (Siam)*, Volume 50.

Tropp, J. A. and Wright, S. J., (2010). Computational methods for sparse solution of linear inverse problems. *Proceedings of the IEEE*, 98(6), pp. 948-958.

Tschumperle, D. and Deriche, R., (2005). Vector-valued image regularization with PDEs: A common framework for different applications. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 27(4), pp. 506-517.

Turk, M. and Pentland, A., (1991). Eigenfaces for recognition. *Journal of cognitive neuroscience*, 3(1), pp. 71-86.

Vandewalle, P., Susstrunk, S. and Vetterli, M., (2006). A frequency domain approach to registration of aliased images with application to super-resolution. *EURASIP Journal on applied signal processing*, Volume 2006, pp. 233-233.

Vuini, E., Gokmen, M. and Groller, E., (2007). Face recognition under varying illumination. *Proceedings WSCG*, pp. 57-64.

Wang, S., Zhang, D., Liang, Y. and Pan, Q., (2012). Semi-coupled dictionary learning with applications to image super-resolution and photo-sketch synthesis. *Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on*, pp. 2216-2223.

Wang, Z. and Bovik, A. C., (2002). A universal image quality index. *IEEE Signal Processing Letters*, Volume 9, pp. 81-84.

Welch, L., (1974). Lower bounds on the maximum cross correlation of signals. *Information Theory, IEEE Transactions on*, 20(3), pp. 397-399.

Xiaoqing, S. L., (2011). Multi-frame Image Super-resolution Reconstruction based on Sparse Representation and POCS. *International Journal of Digital Content Technology and its Applications*, Volume 5(Number 8).

Xie, X., Zheng, W.S., Lai, J. and Yuen, P., (2008). Face illumination normalization on large and small scale features. *Computer Vision and Pattern Recognition (CVPR), IEEE Conference on*, pp. 1-8.

Yang, A., Wright, J., Ma, Y. and Sastry, S., (2007). Feature selection in face recognition: A sparse representation perspective. *submitted to IEEE Transactions Pattern Analysis and Machine Intelligence*.

Yang, J., Wright, J., Huang, T. and Ma, Y., (2008). Image super-resolution as sparse representation of raw image patches. *Computer Vision and Pattern Recognition, CVPR. IEEE Conference on*, pp. 1-8.

Yang, J., Wright, J., Huang, T. and Ma, Y., (2010). Image super-resolution via sparse representation. *Image Processing, IEEE Transactions on*, 19(11), pp. 2861-2873.

Yu, L., Barbot, J., Zheng, G. and Sun, H., (2010). Toeplitz-structured Chaotic Sensing Matrix for Compressive Sensing. pp. 229-233.

Zeyde, R., Protter, M. and Elad, M., (2010). On Single Image Scale-Up using Sparse-Representation. *Springer*, p. pp. 711–730.

Zhang, J., Zhao, C. and Xiong, R., (2012). Image super-resolution via dual-dictionary learning and sparse representation. *Circuits and Systems (ISCAS), IEEE International Symposium on*, pp. 1688-1691.

Zhang, S., Zhao, X. and Lei, B., (2012). Robust facial expression recognition via compressive sensing. *Sensors*, 12(3), pp. 3747-3761.

Zhang, X. and Gao, Y., (2009). Face recognition across pose: A review. *Pattern Recognition*, 42(11), pp. 2876-2896.

Zibetti, M., Bazan, F. and Mayer, J., (2008). Determining the regularization parameters for super-resolution problems. *Signal Processing*, 88(12), pp. 2890-2901.

Zou, W. and Yuen, P., (2010). Very low resolution face recognition problem. *Biometrics: Theory Applications and Systems (BTAS), 4th IEEE International Conference on*, pp. 1-6.

# LIST OF PUBLICATIONS

- **Nadia AL-Hassan**, Sabah A. Jassim, and Harin Sellahewa. "**Enhancing Face Recognition at a Distance using Super Resolution**", 14th ACM Workshop on Multimedia and Security, p.123-132, 2012.

- **Nadia AL-Hassan**, Sabah A. Jassim, and Harin Sellahewa. "**Construction of Dictionaries to Reconstruct High-Resolution Images for Face Recognition**", 6th International Conference on Biometrics, ICB, Madrid, Spain, June 2013.

- **Nadia AL-Hassan**, Harin Sellahewa, and Sabah A. Jassim. "**Super Resolution Based Face Recognition Do we need training image set?**", Proc. SPIE Mobile Multimedia/ Image Processing, Security, and Applications, 8755, 87550 Vol., April 2013.

- Sabah A. Jassim and **Nadia AL-Hassan**. "**Face Recognition in Uncontrolled Condition Can Compressive Sensing and Super-Resolution Meet the Challenge?**", Journal of Biometrics & Biostatistics, 4Vol. (171), P2, 2013.

# APPENDIX

## A Performance of Wavelet-based Face Identification on Super Resolved Image from UBHSD database

| Gallery Set | Probe Set | SR methods | Gallery Image Range | | | |
|---|---|---|---|---|---|---|
| | | | $Range_1$ | $Range_2$ | $Range_3$ | $Range_4$ |
| $SD_{128}$ | $SD_{64}$ | LD-Sp | 66.87 | 65.83 | 65.83 | 62.70 |
| | | $LID_1$ | 66.66 | 65.62 | 65.41 | 62.50 |
| | | GRM | 66.47 | 65.93 | 65.03 | 62.70 |
| | | TCRM | 66.87 | 65.83 | 65.63 | 62.70 |
| | | $LID_2$ | 66.87 | 65.62 | 65.83 | 62.50 |
| | | ROM | 66.66 | 65.62 | 65.62 | 62.70 |
| | | BC | 66.87 | 65.62 | 65.83 | 62.50 |
| $SD_{64}$ | $HD_{128}$ | LD-Sp | 68.12 | 62.70 | 61.45 | 61.87 |
| | | $LID_1$ | 68.33 | 62.70 | 61.45 | 61.87 |
| | | GRM | 68.12 | 62.70 | 61.45 | 61.87 |
| | | TCRM | 68.33 | 62.70 | 61.45 | 61.87 |
| | | $LID_2$ | 68.33 | 62.70 | 61.45 | 61.87 |
| | | ROM | 68.12 | 62.70 | 61.45 | 61.87 |
| | | BC | 68.33 | 62.72 | 61.40 | 61.87 |
| $HD_{128}$ | $SD_{64}$ | LD-Sp | 59.58 | 62.29 | 64.58 | 66.87 |
| | | $LID_1$ | 59.37 | 62.29 | 64.37 | 66.45 |
| | | GRM | 59.46 | 62.50 | 64.79 | 66.66 |
| | | TCRM | 59.58 | 62.50 | 64.58 | 66.66 |
| | | $LID_2$ | 59.37 | 62.50 | 64.79 | 66.45 |
| | | ROM | 59.37 | 62.29 | 64.37 | 66.45 |
| | | BC | 59.58 | 62.50 | 64.58 | 66.66 |
| $SD_{64}$ | $SD_{64}$ | LD-Sp | 67.08 | 65.62 | 65.41 | 63.12 |
| | | $LID_1$ | 66.66 | 65.62 | 65.20 | 62.91 |
| | | GRM | 67.08 | 66.04 | 65.41 | 63.12 |
| | | TCRM | 67.08 | 68.83 | 65.20 | 63.12 |
| | | $LID_2$ | 67.08 | 68.04 | 65.20 | 63.12 |
| | | ROM | 66.87 | 65.41 | 65.41 | 62.70 |
| | | BC | 66.87 | 68.04 | 65.20 | 62.70 |

**Table A.1:** Recognition accuracy rates (%) for the UBHSD database using different SR dictionaries to super resolve Full--face image, based on $LL_3$ subband.

| Gallery Set | Probe Set | SR methods | Gallery Image Range | | | |
|---|---|---|---|---|---|---|
| | | | $Range_1$ | $Range_2$ | $Range_3$ | $Range_4$ |
| $SD_{128}$ | $SD_{64}$ | LD-Sp | 69.37 | 69.16 | 69.79 | 58.75 |
| | | $LID_1$ | 69.37 | 69.16 | 69.58 | 58.54 |
| | | GRM | 69.16 | 69.37 | 69.58 | 58.54 |
| | | TCRM | 69.58 | 69.37 | 69.79 | 58.95 |
| | | $LID_2$ | 69.37 | 69.16 | 69.58 | 58.95 |
| | | ROM | 69.58 | 69.16 | 69.37 | 58.75 |
| | | BC | 69.16 | 69.16 | 69.58 | 58.54 |
| $SD_{64}$ | $HD_{128}$ | LD-Sp | 69.37 | 63.33 | 73.33 | 61.45 |
| | | $LID_1$ | 68.95 | 63.54 | 73.95 | 61.04 |
| | | GRM | 69.16 | 63.95 | 74.16 | 61.04 |
| | | TCRM | 69.16 | 63.54 | 74.16 | 61.04 |
| | | $LID_2$ | 69.37 | 63.33 | 73.95 | 61.45 |
| | | ROM | 69.16 | 63.54 | 74.58 | 61.25 |
| | | BC | 69.37 | 63.33 | 73.33 | 61.45 |
| $HD_{128}$ | $SD_{64}$ | LD-Sp | 65.41 | 69.58 | 64.58 | 68.95 |
| | | $LID_1$ | 65.83 | 69.58 | 64.79 | 69.16 |
| | | GRM | 65.04 | 69.75 | 64.56 | 69.16 |
| | | TCRM | 65.83 | 70.20 | 64.79 | 69.37 |
| | | $LID_2$ | 65.83 | 69.58 | 64.79 | 68.95 |
| | | ROM | 65.41 | 69.58 | 64.79 | 68.95 |
| | | BC | 65.41 | 69.58 | 64.58 | 68.95 |
| $SD_{64}$ | $SD_{64}$ | LD-Sp | 69.58 | 68.12 | 69.79 | 60 |
| | | $LID_1$ | 69.58 | 68.54 | 69.58 | 59.16 |
| | | GRM | 69.58 | 68.12 | 69.79 | 60 |
| | | TCRM | 69.79 | 68.54 | 69.79 | 59.90 |
| | | $LID_2$ | 69.58 | 68.12 | 69.58 | 60 |
| | | ROM | 69.16 | 68.33 | 69.16 | 59.37 |
| | | BC | 69.79 | 68.12 | 69.58 | 59.37 |

**Table A.2:** Recognition accuracy rates (%) for the UBHSD database using different SR dictionaries to super resolve Full--face image, based on $HL_3$ subband.

| Gallery Set | Probe Set | SR methods | Gallery Image Range | | | |
|---|---|---|---|---|---|---|
| | | | $Range_1$ | $Range_2$ | $Range_3$ | $Range_4$ |
| $SD_{128}$ | $SD_{64}$ | LD-Sp | 53.95 | 48.75 | 46.87 | 45.20 |
| | | $LID_1$ | 53.12 | 49.37 | 46.66 | 45.41 |
| | | GRM | 53.95 | 49.58 | 46.87 | 45 |
| | | TCRM | 54.16 | 50 | 46.87 | 49.20 |
| | | $LID_2$ | 53.12 | 49.37 | 46.66 | 45 |
| | | ROM | 53.54 | 49.58 | 46.45 | 45.41 |
| | | BC | 53.95 | 48.75 | 46.66 | 45.20 |
| $SD_{64}$ | $HD_{128}$ | LD-Sp | 57.91 | 47.50 | 44.79 | 40.62 |
| | | $LID_1$ | 57.91 | 48.75 | 45.41 | 41.04 |
| | | GRM | 57.50 | 47.08 | 46.04 | 40 |
| | | TCRM | 57.29 | 48.12 | 45.20 | 40.62 |
| | | $LID_2$ | 57.91 | 47.50 | 44.79 | 41.04 |
| | | ROM | 57.50 | 47.29 | 45.83 | 41.25 |
| | | BC | 57.91 | 48.75 | 46.04 | 40.62 |
| $HD_{128}$ | $SD_{64}$ | LD-Sp | 52.70 | 50.20 | 49.79 | 47.91 |
| | | $LID_1$ | 53.54 | 50.62 | 50 | 47.91 |
| | | GRM | 53.33 | 50.41 | 50.03 | 48.83 |
| | | TCRM | 53.12 | 51.04 | 50 | 48.54 |
| | | $LID_2$ | 53.54 | 51.04 | 50 | 48.54 |
| | | ROM | 52.91 | 50.62 | 50.20 | 48.12 |
| | | BC | 53.54 | 50.41 | 50 | 47.91 |
| $SD_{64}$ | $SD_{64}$ | LD-Sp | 54.58 | 48.95 | 46.66 | 42.91 |
| | | $LID_1$ | 53.95 | 49.58 | 46.87 | 42.91 |
| | | GRM | 54.79 | 49.79 | 46.87 | 43.54 |
| | | TCRM | 69.79 | 68.54 | 69.79 | 59.90 |
| | | $LID_2$ | 53.95 | 48.95 | 46.87 | 43.54 |
| | | ROM | 54.58 | 49.37 | 46.87 | 44.58 |
| | | BC | 54.79 | 49.79 | 46.87 | 42.91 |

**Table A.3:** Recognition accuracy rates (%) for the UBHSD database using different SR dictionaries to super resolve Full--face image, based on $HH_3$ subband.

# B Performance of Face Identification on Super Resolved Feature Image from UBHSD database

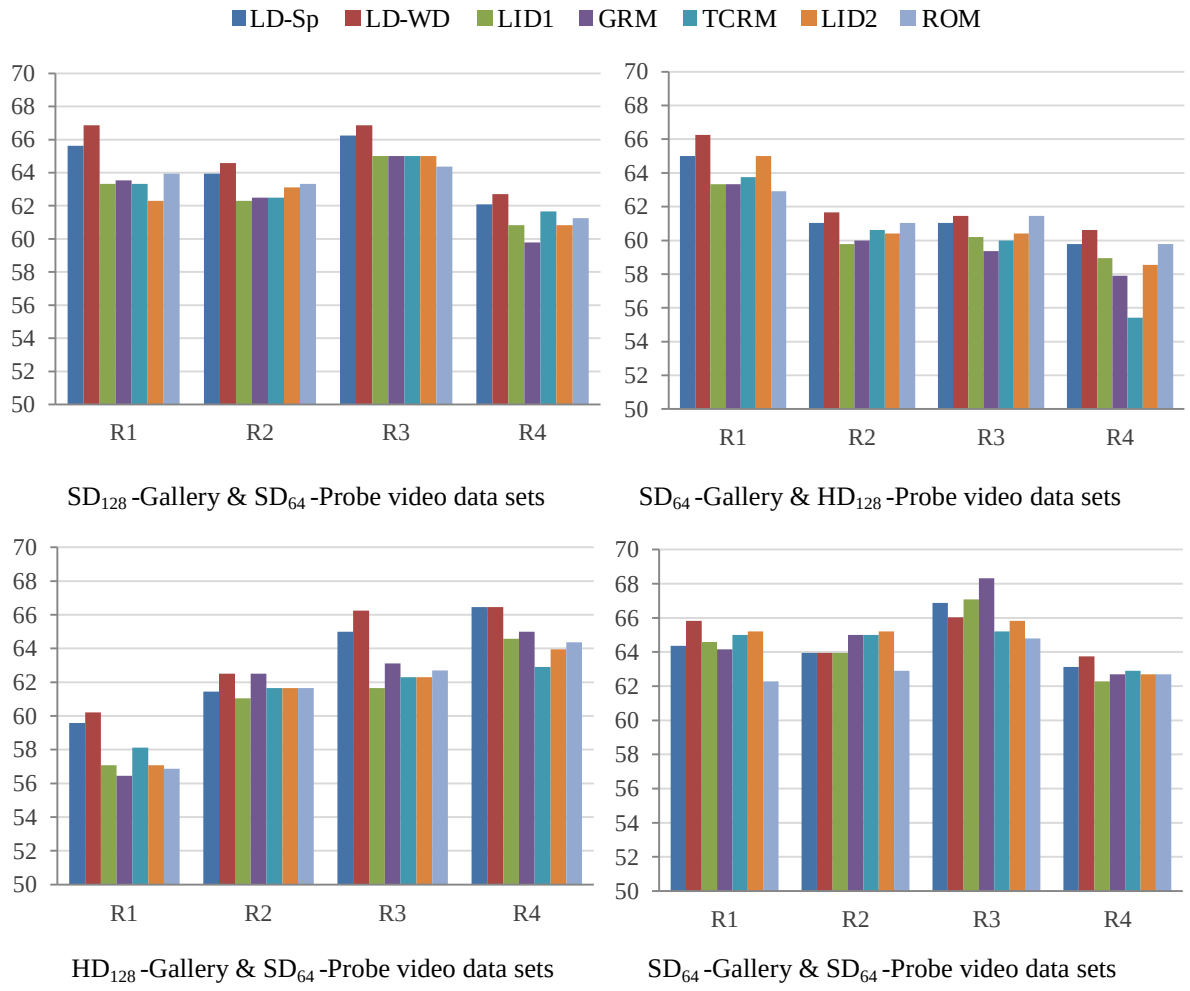



**Figure B.1:** Comparisons between different SR dictionaries that used for super resolve $LL_3$ wavelet subband of the LR image from UBHSD data.

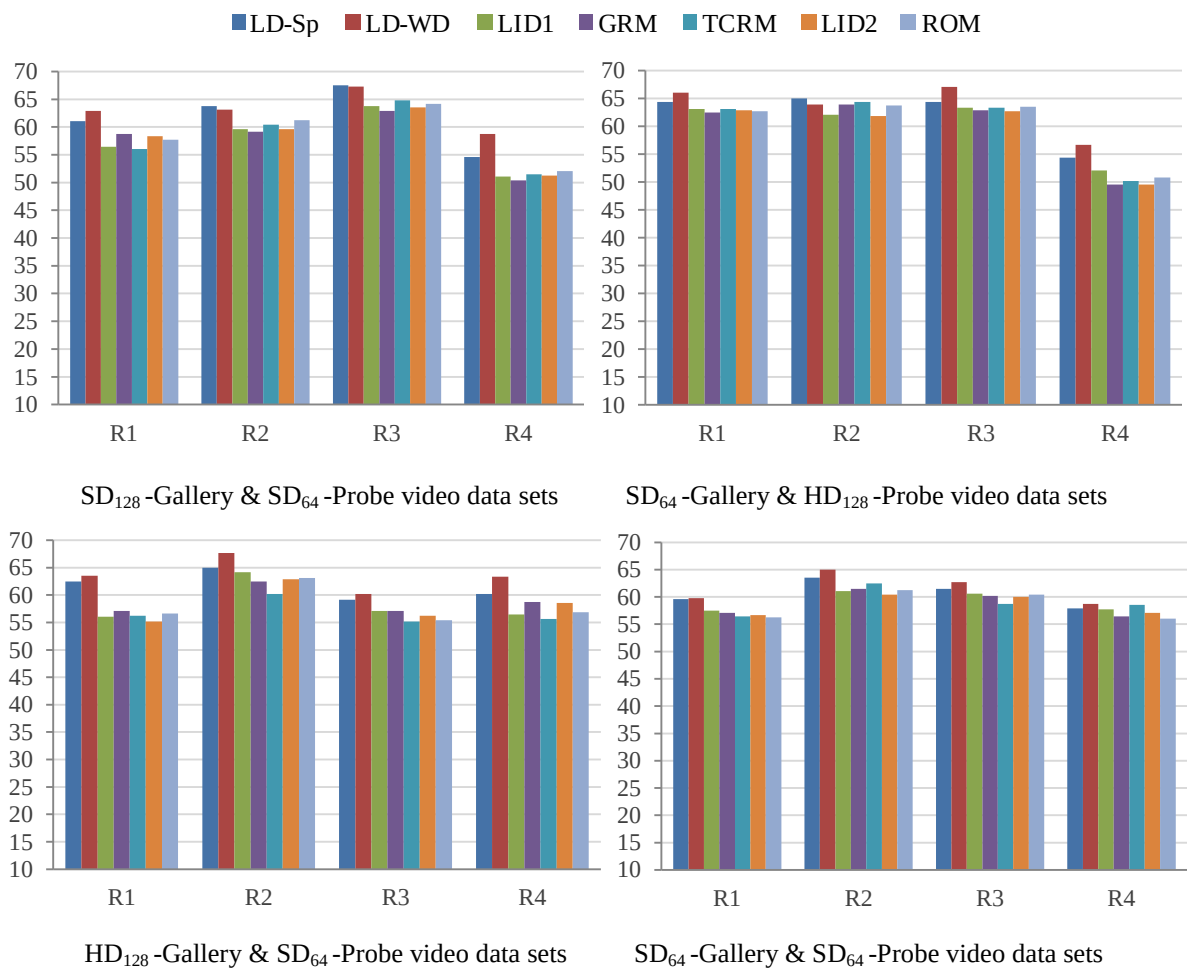



**Figure B.2:** Comparisons between different SR dictionaries that used for super resolve $HL_3$ wavelet subband of the LR image from UBHSD data.

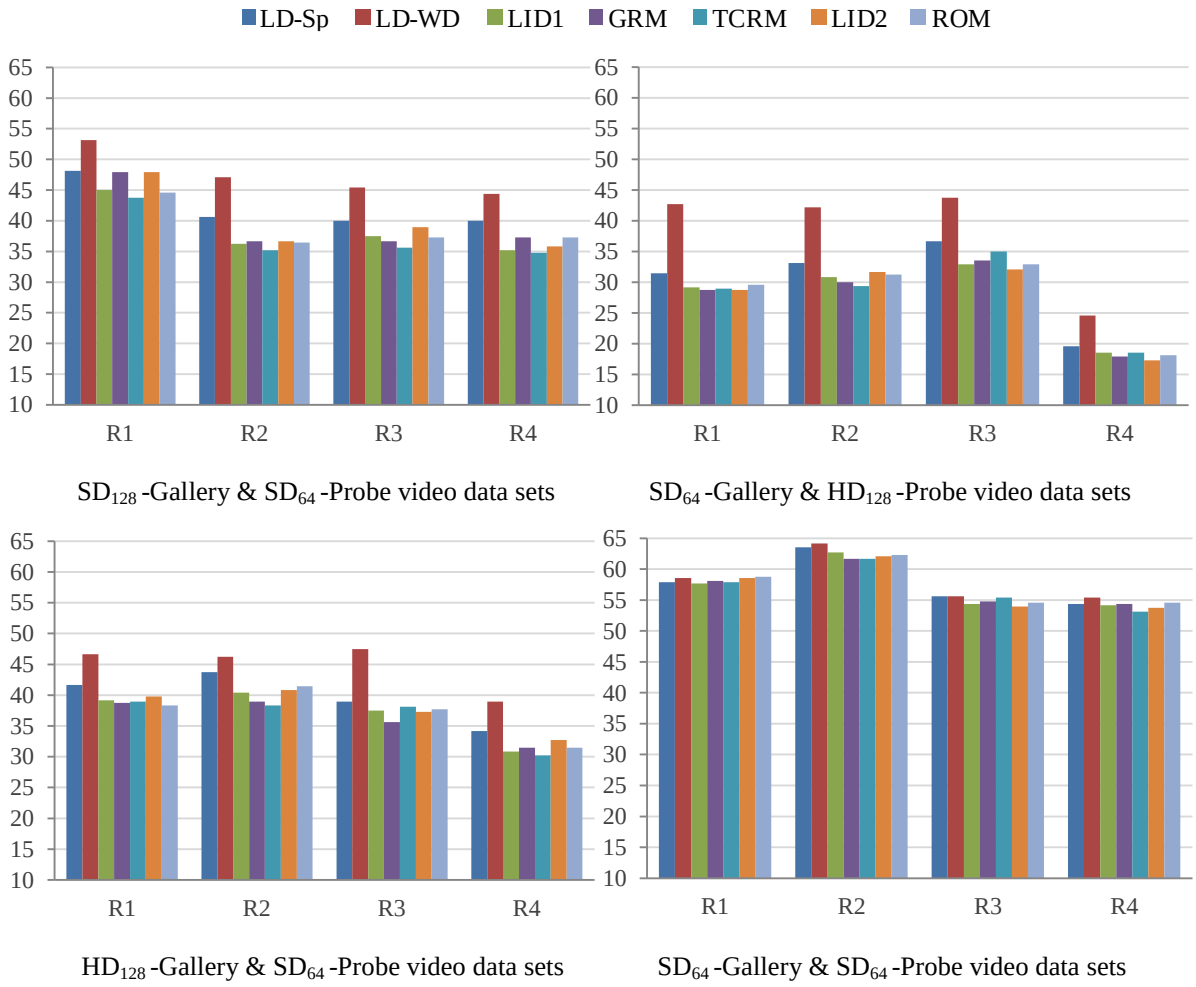



**Figure B.3:** Comparisons between different SR dictionaries that used for super resolve $HH_3$ wavelet subband of the LR image from UBHSD data.