REPUBLIC OF TURKEY
YILDIZ TECHNICAL UNIVERSITY
GRADUATE SCHOOL OF NATURAL AND APPLIED SCIENCES

ROBUST CLASSIFICATION BASED ON SPARSITY

ELENA BATTINI SÖNMEZ

THESIS FOR THE DEGREE OF DOCTOR OF PHILOSOPHY
COMPUTER ENGINEERING DEPARTMENT

ADVISOR
ASSISTANT PROF. DR. SONGÜL ALBAYRAK

ISTANBUL, 2011

REPUBLIC OF TURKEY
YILDIZ TECHNICAL UNIVERSITY
GRADUATE SCHOOL OF NATURAL AND APPLIED SCIENCES

ROBUST CLASSIFICATION BASED ON SPARSITY

Presented by Elena BATTINI SÖNMEZ on 22.12.2011 and accepted by the following committee on behalf of the Computer Engineering Department, Yıldız Technical University, in partial satisfaction of the requirements for the **DEGREE of DOCTOR of PHILOSOPHY.**

Thesis's Advisor

Assist. Prof. Dr. Songül Albayrak

Yıldız Technical University

Jury's Members

Assist. Prof. Dr. Songül Albayrak

Yıldız Technical University ________________________

Prof. Dr. Bülent SANKUR

Boğaziçi University ________________________

Prof. Dr. Hasan Dağ

Kadır Has University ________________________

Prof. Dr. A. Çoşkun Sönmez

Yıldız Technical University ________________________

Assist. Prof. Dr. Güneş Karabulut Kurt

Istanbul Technical University ________________________

ACKNOWLEDGEMENTS

This thesis is the result of several years of work, impossible to accomplish without the help of many people. I take this opportunity to thank all of them.

I thank very much my advisor, Assist. Prof. Songül Albayrak, for being always present at my side, and giving me precious advice as well as the moral to overcome difficulties. It would have been very difficult to complete this research without her encouragement.

I thank very much Prof. Bülent Sankur for the big effort in teaching and directing me in this research. I am very grateful for the enormous amount of time spent with me discussing different problems ranging from theoretical issues down to technical details.

My special thanks to my family, who kept the joy in my life.

Finally, I would like to thank the family of my husband, for the never ending help, all my friends for being always nearby me, and all of my colleagues at Istanbul Bilgi University for their patient understating of my overwhelming schedule.

August, 2011

Elena Battini Sönmez

CONTENTS

Page

LIST OF SYMBOLS vii

LIST OF ABBREVIATIONS viii

LIST OF FIGURES x

LIST OF TABLES xii

ABSTRACT xiii

ÖZET xiv

CHAPTER 1

INTRODUCTION

1.1 Literature Review 1
1.2 Aims of the Thesis 5
1.3 Hypothesis 6

CHAPTER 2

THE SPARSE REPRESENTATION BASED CLASSIFIER (SRC) 7

2.1 Sparse Representation of Signals 7
2.1.1 Synthesis versus Analysis 9
2.1.2 Compressive Sensing Theory 10
2.1.3 Under vs Over-Determined System of Equations 14
2.1.4 Synthesis Algorithms 15
2.2 Sparse Representation based Classifier 16
2.2.1 A case study: Sparse Representation for Face Recognition 19
2.2.2 Algorithms 20

CHAPTER 3

STUDY OF THE CHARACTERISTICS OF SRC 28

3.1 Critical Parameters 28
  3.1.1 Sparsity Level 28
  3.1.2 Coherence of the Dictionary 30
  3.1.3 Mutual Coherence Dictionary-Test Sample 30
  3.1.4 Confidence Level 31
3.2 Statistical Studies of SRC 32
  3.2.1 Coefficients Correlation Matrix 32
  3.2.2 Coefficients Histogram 34

CHAPTER 4

SRC WITH CORRUPTED FACES 36

4.1 Classification of Corrupted Faces via SRC 36
4.2 Description of the Database 37
4.3 Experiments 37
  4.3.1 Down-Sampling Subspaces 39
  4.3.2 Effect of the Degree of Over-completeness 40
  4.3.3 Noise 42
  4.3.4 Illumination Compensation Capabilities 43
  4.3.5 Geometric Distortion: shift, in-plane rotation and scaling 45

CHAPTER 5

GEOMETRIC NORMALIZATION AND BLOCK-BASED APPROACH 52

5.1 Geometric Normalization 52
5.2 Block Based Classification 59
  5.2.1 Fusing Algorithms 61

CHAPTER 6

SRC WITH EMOTIONAL FACES 64

6.1 Classification of Emotional Faces via SRC 65
6.2 Description of the Database 65
6.3 Experiments 66
  6.3.1 Emotion Classification 66
  6.3.2 Action Units Identification 71
  6.3.3 Identification Despite Emotions 74

CHAPTER 7

CONCLUSIONS AND FUTURE WORK 78

7.1 Conclusions and Future Work 78

REFERENCES 83

CURRICULUM VITAE 87

LIST OF SYMBOLS

C	number of classes
class$_i$	a general class
$D \in \Re^{N \times M}$	dictionary
DictCoh	coherence of the dictionary (D)
DictTestCoh	mutual coherence between dictionary and test sample
m_i	cardinality of class$_i$
$S \in \Re^{M \times M}$	sparsifying matrix
subject$_i$	a general subject
SRC_miss	number of samples mis-classified by the SRC algorithm
$x \in \Re^{M}$	a coefficients' vector
$X \in \Re^{M \times K}$	matrix of coefficients' vector
$y \in \Re^{N}$	a test sample, the signal under observation
$Y \in \Re^{N \times K}$	matrix of test samples

LIST OF ABBREVIATIONS

2D	two dimensional
3D	three dimensional
ALM	Augmented Lagrange Multiplier
An	Angry
AU	Action Unit
BP	Basis Pursuit
CL	Compressibility Level
Co	Contempt
CS	Compressive Sensing or Compressive Sampling
CF	Cropped Faces
CFGN	Cropped Faces with Geometric Normalization
CMC	Cumulative Match Count
DCT	Discrete Cosine Transform
DFS	Distance from Face Space
DFT	Discrete Fourier Transform
Di	Disgust
DoG	Difference of Gaussian
DWT	Discrete Wavelet Transform
FACS	Facial Action Coding System
FRGC	Face Recognition Grand Challenge
Fe	Fear
FL	Face Landmark
Ha	Happy
HistEq	Histogram Equalization
IOD	Inter Ocular Distance
LDA	Linear Discriminant Analysis
LOSO	Leave One Subject Out
MCC	Mean of the Class Coefficients
MP	Matching Pursuit
NC	Normalization Coefficient
OMP	Orthogonal Matching Pursuit
PCA	Principal Component Analysis
RankN	Rank Normalization
RIP	Restricted Isometry Property

Sa	Sadness
SL	Sparsity Level
SRC	Sparse Representation-based Classifier
Su	Surprise
SVM	Support Vector Machine
W	Window

LIST OF FIGURES

Page

Figure 2.1 The synthesis or analysis problem 8
Figure 2.2 The synthesis or analysis problem (matrix wise) 9
Figure 2.3 Sparse versus compressible signals 11
Figure 2.4 L2 versus L1 balls 13
Figure 2.5 The cross-and-bouquet model for face recognition 18
Figure 3.1 Residuals of a correct classified test sample 31
Figure 3.2 Residuals of a misclassified test sample 31
Figure 3.3 Correlation matrix of the 30 non-ordered coefficients of target class 33
Figure 3.4 Correlation matrix of the 30 coefficients of the target class ordered by their algebraic values 33
Figure 3.5 Correlation matrix of the 30 coefficients of the target class ordered by their absolute values 34
Figure 3.6 Histograms of target coefficients no.1, 2, 3, 4, 29, and 30 35
Figure 4.1 Elevation and azimuth angles 37
Figure 4.2 Samples of faces under adverse conditions 38
Figure4.3 Effects of the gallery size on classifiers’ performance 41
Figure 4.4 Performance (%) under Gaussian noise (plot of Table 4.2, left side) 43
Figure 4.5 Performance (%) under salt&pepper noise (plot Table 4.2, right side) 43
Figure 4.6 Effects of the shift geometric distortion (plot of Table 4.6, left side) 47
Figure 4.7 Effects of the rotation geometric distortion (plot of Table 4.6, center) 47
Figure 4.8 Effects of the zoom geometric distortion (plot of Table 4.6, right side) 48
Figure 5.1 An original and a cropped face in the CK+ database 53
Figure 5.2 Cropped Faces 53
Figure 5.3 Geometric normalization steps 54
Figure 5.4 Original and rotated face 54
Figure 5.5 Rotated and zoomed face 55
Figure 5.6 Photometric normalization 55
Figure 5.7 Aligned faces cropped with a fix size window 56
Figure 5.8 Aligned faces cropped with face landmarks 56
Figure 5.9 A face image partitioned into 2×2 and 3×3 blocks 59
Figure 5.10 Snapshot of a 3D dictionary 59
Figure 5.11 Conversion from a 3D to a 2D matrix of scores 60
Figure 5.12 Masks used for the fusing step 61
Figure 5.13 Rank based algorithm 63
Figure 6.1 The emotion classification problem 67

Figure 6.2 Emotional face with multiple AUs 72
Figure 6.3 CMC scores for AUs identification on the entire face 73
Figure 6.4 CMC scores at N+5 for AU identification. Bars left to right: whole face, average 3×3 blocks, upper/lower AUs 3×3 blocks 74

LIST OF TABLES

Page

Table 2.1 Pseudo-code of DFS 20
Table 2.2 Pseudo-code of MCC 21
Table 2.3 Pseudo-code of biggest coefficient DFS 22
Table 2.4 Pseudo-code of three biggest coefficients DFS 23
Table 2.5 Pseudo-code of minimum LS norm 24
Table 2.6 Pseudo-code of different cardinality class DFS1 25
Table 2.7 Pseudo-code of different cardinality class DFS2 26
Table 2.8 Pseudo-code of MCC with absolute value 27
Table 4.1 Effects of down-sampling subspaces and image resolution on classifiers' performance (%) 40
Table 4.2 Effects of the gallery size on classifiers' performance 41
Table 4.3 Performance (%) under Gaussian (left) and salt&pepper noise (right) 42
Table 4.4 Training and test sets of experiments on illumination 44
Table 4.5 Illumination compensation capability 44
Table 4.6 Blind experiments on shift (left), rotation (centre), zoom (right) 47
Table 4.7 Blind experiment on shift 49
Table 4.8 Informed experiment on shift 50
Table 5.1 Performance of different experimental setups 57
Table 6.1 Confusion matrix of holistic SRC with CF (benchmark results) 69
Table 6.2 Confusion matrix of holistic SRC with CFGN (benchmark results 70
Table 6.3 Confusion matrix for 2×2 blocks based SRC on CF (benchmark results) 71
Table 6.4 Population of selected AUs 72
Table 6.5 Distribution of the number of subjects sharing emotion pairs 75
Table 6.6 Subject recognition rate (percentage) across expression 76
Table 6.7 Subject recognition rate (percentage) with a richer dictionary 77

ABSTRACT

ROBUST CLASSIFICATION BASED ON SPARSITY

Elena BATTINI SÖNMEZ

Computer Engineering Department

PhD Thesis

Advisor: Assist. Prof. Dr. Songül ALBAYRAK

Classification is one of the most important problems in machine learning with a range of applications, including computer vision. Given training images from different classes, the problem is to find the class that a test sample belongs to. This thesis focuses on 2D face classification under adverse conditions. The problem is particularly hard in the presence of disturbances such as variations in illumination, expressions, pose, alignment, occlusion and resolution. Despite great interest in the past years, current pattern recognition methods still fail to classify faces in the presence of all types of variations. Recent developments in the theory of compressive sensing have inspired a sparsity based classification algorithm, which turns out to be very successful. This study investigates the potentialities of the Sparse Representation based Classifier (SRC) and, in parallel, it monitors the behaviour of some factors, which can reflect its performance. All experiments use the Extended Yale B and the Extended Cohn Kanade databases. The first dataset stores images with changes in illumination and has a cropped sub-directory of aligned faces, which allows for inserting a controlled amount of misalignment. The second database has coded action units and emotions, which permits to challenge both action units and emotion classification problems as well as the identity recognition despite emotions issue. Experimental results place SRC into the shortlist of the most successful classifiers mainly because of its inner robustness and simplicity.

Key Words: Classification, sparsity, face recognition, action unit, emotion, incoherence

---

YILDIZ TECHNICAL UNIVERSITY

GRADUATE SCHOOL OF NATURAL AND APPLIED SCIENCES

ÖZET

SEYREKLİĞE DAYALI GÜRBÜZ SINIFLAYICI

Elena BATTINI SÖNMEZ

Bilgisayar Mühendisliği Bölümü
Doktora Tezi

Tez Danışmanı: Yrd. Doç. Dr. Songül ALBAYRAK

Bilgisayarla görmenin de içinde olduğu farklı uygulamalarda kullanılan sınıflama konusu makine öğrenmesindeki en önemli problemlerden biridir. Sınıflama, farklı sınıflara ait eğitim örneklerinin sisteme verilerek test örneğinin hangi sınıfa ait olduğunun bulunması problemidir. Bu tez çalışmasında aydınlatma farklılıkları, yüz ifadesi, poz, yanlış hizalama ve çözünürlük gibi bozucu etkiler altında 2B yüz sınıflama problemine odaklandık. Son yıllarda yüz tanıma konusuna çok büyük bir ilgi olmasına karşın, mevcut örüntü tanıma metodları tüm bu bozucu etkilerin var olduğu durumlarda yüz sınıflamada başarısız olmaktadır. Sıkıştırmalı algılama teorisinde görülen gelişmeler, seyreklik tabanlı sınıflayıcılara doğru yayılmış ve başarılı sonuçlar alınmıştır. Bu çalışmada seyrek yaklaşım tabanlı sınıflama algoritmasının potansiyeli ve başarımını etkileyen faktörlere karşı davranışı araştırılmıştır. Çalışmada gerçekleştirilen tüm deneylerde Exteded Yale B ve Extended Cohn Kanade veritabanları kullanılmıştır. Birincisinde aydınlatma farklılıkları olan görüntüler kontrol edilebilir miktarda hizalamaya izin verecek şekilde kırpılarak alt dizinlere yerleştirilmiştir, eylem birini ve yüzdeki duygunun kodlandığı ikinci veritabanı sınıflama problemi kadar yüz ifadesine rağmen kişi tanımaya da uygundur. Deneylerdeki sonuçlar SRC'yi, başta gürbüzlüğü ve basitliğinden dolayı en başarılı sınıflandırıcılar listesine sokmuştur.

Anahtar Kelimeler: Sınıflama, seyreklik, yüz tanıma, eylem birini, duygu, uyumsuzluk

---

CHAPTER 1

INTRODUCTION

1.1 Literature Review

Automatic identification and verification of humans using facial information has been one of the most active research areas in computer vision. The interest on face recognition is fuelled by the identification requirements for access control and for surveillance tasks, whether as a means to increase work efficiency and/or for security reasons. Face recognition is also seen as an important part of next-generation smart environments [1], [2].

Face recognition algorithms under controlled conditions have achieved reasonably high levels of accuracy. However, under non-ideal, uncontrolled conditions, as often occur in real life, their performance becomes poor. Their main handicaps are the changes in face appearances caused by such factors as occlusion, illumination, expression, misalignment, pose, make-ups and aging. In fact the intra-individual face differences due to some of these disturbances can easily be larger than the inter-individual variability [3], [4], [5].

We briefly point out below some of the main roadblocks to wide scale deployment of reliable face biometry technology.

Effects of Illumination: changes in illumination can vary the overall magnitude of the light intensity reflected back from an object and modify the pattern of shading and shadows visible in an image [5]. It was shown that varying illumination is the most detrimental to both human and machine accuracies in recognizing faces. It suffices to quote simply the fact that in the Face Recognition Grand Challenge (FRGC) the 17 algorithms competing in the controlled illumination track have achieved a median verification rate of 0.91 while in contrast, the 7 algorithms competing in the uncontrolled illumination experiment have achieved a median verification rate of only 0.42 (both figures at a false acceptance rate of 0.001). The difficulties posed by variable illumination conditions, therefore, remain still one of the main roadblocks to reliable face recognition systems.

Effects of Expression: Facial expression is known to affect the face recognition accuracy though in the current literature, a full-fledged analysis of the deterioration caused by expressions has not been documented. Instead, most studies either focus on expression recognition alone or on face identification alone. It is quite interesting that this dichotomy is also encountered in biological vision. There is strong evidence that facial identity and expression might be processed by separate systems in the brain, or at best they are loosely linked [6]. To state the same concept in a different way, faces with emotions present two interesting and concomitant problems:

- The first one is automatic understanding of the mental and emotional state of the subject, as reflected on the facial expression. It is stated that about 55% of interpersonal feelings and attitudes, such as like and dislike, can be conveyed via facial expressions. In other words, the human face is a rich source of non-verbal information providing clues to understand social feelings and facilitating the interpretation of intended messages.
- The second problem is the identity recognition based on emotional faces. The plethora of publications on face biometry has indicated that the face remains the holy grail of biometric recognition. Unfortunately, while face enables remote biometry, it is handicapped by several factors among which there is the presence of expressions, [7].

Effects of Misalignment: Real-world images have faces with background, which requires the use of face detectors to locate and crop the face before being able to classify it. In other words, in an automatic face recognition system, face detection and cropping are necessary pre-processing steps and its final recognition rate heavily relies on their accuracies. At the current state of the art, automatic cropping or even manual face cropping may result in big misalignments, such as faces with translation, rotation and scaling, which, in turn, decrease the success rate of the face recognition system. In 2005 Gong et al. [8] proposed a new algorithm to increase the robustness of the classical Gabor features to misalignment, in 2009 Wagner et al. [9] introduced an iterative sparse classifier, which is robust to shifts and rotations, in 2010, the same problem is faced by Yan et al. [10] who proposed a general L1 norm minimization formulation. Nevertheless, bad cropping and misalignment are still among the main factors impeding accurate face recognition.

Effects of Pose: Facial pose or viewing angle is a major impediment to machine-based face recognition. As the camera pose changes, the appearance of the face changes due to projective deformation (causing stretching and foreshortening of different part of face); also self-occlusions and/or uncovering of face parts can arise, [11]. The resulting effect is that image-level differences between two views of the same face are much larger than those between two different faces viewed at the same angle. While machines fail badly in face recognition under viewing angle changes, that is, when trained on a gallery of a given pose and tested with a probe of set of a different viewing angle, humans have no difficulty in recognizing faces at arbitrary poses. It has been reported in [11] that the performance of Principal Component Analysis (PCA) based method decreases dramatically beyond 32 degree yaw and those for Linear Discriminant Analysis (LDA) beyond 17 degree rotation.

Effects of Occlusion: The face may be occluded by facial accessories such as sunglasses, a snow cap, a scarf, by facial hair or other paraphernalia. Furthermore, the subjects in an effort to eschew being identified can purposefully cover parts of their face. Although it is very difficult to systematically experiment with all sorts of natural or intentional occlusions, results reported in [11] show that methods like PCA and LDA fail quite badly (e.g., sunglasses and scarf scenes in AR database). Recent work on recognition by parts shows that methods that rely on local information, can perform fairly well under occlusion [12], [13], [14].

Effects of Low Resolution: The performance loss of face recognition with decreasing resolution is well known and documented, [15]. For example, 30 percentage point drops are reported in [15] as the resolution changes from 65×65 to 32×32 pixel faces.

The robustness to varying resolution becomes relevant especially in uncontrolled environments where the face may be captured by a security camera at various distances within its field of view.

In recent years the growing attention in the study of the compressive sensing theory fuelled an increase interest in the sparse representation of signals which has suggested a different approach to achieve robust classification. The basic idea is to cast classification as a sparse representation problem, utilizing new mathematical tools from the compressive sampling studies [16], [17], [18]. The new algorithm was first introduced by Wright et al. in [14]; it proposes a new classification technique called Sparse Representation-based Classifier (SRC) because based on the sparse representation of the test sample. The novelty of this algorithm is in the classification method, which uses the coefficient vector, x, returned by a synthesis algorithm claiming that x has enough discriminative power to classify the current test sample. It is important to point out that, instead of the classical dictionaries, such as Fourier, DCT, or Wavelets, Wright et al. used an ad-hoc dictionary, D, built by aligning, column after column, the down-sampled training images themselves. The main idea underlying this approach is that, if sufficient training samples are available for each class, then it is possible to represent every test sample as a linear combination of just those training samples belonging to the same class. Seeking the sparsest representation automatically discriminates between the various classes present in the training database and also provides a simple and effective mechanism for rejecting outliers, invalid test samples not belonging to any class of the training set. SRC method was inspired by the compressive sensing theory and it works very well even if, at first glance, it violates some of its important requirements. One of the aims of this study is to distinguish among the relevant and the irrelevant requirements of SRC.

1.2 Aims of the Thesis

The aim of this work is to explore the capabilities of the sparse representation based classifier (SRC). In order to do that, we tackled different classification issues using two databases of faces; that is, we worked on identity recognition with faces corrupted by noise, geometric distortions or illumination changes, as well as we challenged the emotion and action unit identification problems and we studied the deterioration caused by expressions on the identity recognition task with emotional faces. While running all experiments, we monitored different parameters considered critical for the performance of SRC, such as the sparsity level (SL) of the coefficients' vector and the level of incoherence of the dictionary (*DictCoh*) so as to detect "if" and "how" those measurements are really correlated to the performance of SRC, and we introduced other critical parameters capable of predict the success rate of SRC. Moreover, we made statistical studies on the correlation matrix of the target coefficients and on the distribution of the coefficients vector with the aim to understand and justify the performance of SRC. Finally, for some critical experiments, we improved the original success rate of SRC by enlarging its dictionary, using a block-based approach or pre-processing all samples with geometric normalization.

Chapter 2 gives a general overview of the sparse representation and compressive sensing theory, followed by an introduction of the sparse representation based classifier (SRC), with a case study on face recognition. Chapter 3 introduces some parameters, which are considered critical for the performance of SRC, it proposes new important measurements related to its success rate, and it stores the results of statistical studies on the distribution of the coefficients' vector and their correlation matrix. Chapters 4, 5 and 6 are empirical chapters. The aim of all experiments is to test the robustness of the sparse classifier to adverse conditions, to identify the critical parameters correlated to the performance of SRC, and to explore possible variations. More in detail, Chapter 4 works with a database of clean aligned and cropped faces affected by changes in illumination. This database naturally allows us to study the robustness of SRC to illumination changes as well as it permits to insert a controlled amount of shit-rotation-zoom or unwanted information and to run experiments on in-plane geometric distortion and noise. For some relevant cases, Chapter 4 monitors also the critical parameters given in Chapter 3; some of the experiments documented in Chapter 4 were previously presented in [19] and [20]. Chapter 5 details the geometric normalization pre-processing technique together with the block based approach. Chapter 6 works with a database of coded emotional faces, it presents experimental results on emotion and action unit identifications as well as it investigates the deterioration caused by the presence of expressions in the identity recognition task. Conclusions are drawn in Chapter 7.

All experiments were run in the MATLAB R2007b environment on different platforms.

1.3 Hypothesis

We want to prove the conjecture that sparse representation enables the creation of templates robust against various factors impeding accurate face recognition such as illumination, expression, misalignment, and noise. We demonstrate, experimentally, the robustness of the SRC classifier under adverse conditions. Important to notice that the idea of the SRC algorithm was inspired by the compressive sensing (CS) theory, but then, on the practical side, SRC violates the fundamental theoretical assumption of CS such as the restricted isometry property (RIP), which is measured as the level of incoherence of the dictionary. Nevertheless, it is a very successful classification method, inherently robust due to the presence of the L1 norm. This study points out the differences between CS and SRC, and conjectures that compressibility is the only pre-requisite necessary for successful classification.

CHAPTER 2

THE SPARSE REPRESENTATION BASED CLASSIFIER (SRC)

2.1 Sparse Representation of Signals

A sparse representation of signals allows for a high level representation of the data, which is both the most natural and insightful one. In the signal processing community, looking for a compact representation of the data is equivalent to search for an orthogonal transformation where the signal can be represented, in the new domain, as a superposition of few basic elements or atoms; classical examples of such a basis are Discrete Fourier Transform (DFT) with its variation Discrete Cosine Transform (DCT), and the Discrete Wavelet Transform (DWT). The usual approach is to choose one orthogonal basis and use it to represent signals, because the basis is a standard one, the transformation does not consider the particularities of the signals.

A more recent approach rejects orthogonality and standard basis on behalf of over-complete dictionaries. A dictionary, D, is a collection of parameterized waveforms where each waveform, d_i, is a discrete time signal of length N called atom and it has unit length, $\| d_i \| = 1$. Dictionaries are complete, if they contain exactly N linearly independent atoms, and over-complete, if they contain more than N atoms; D is an over-complete dictionary. The use of over-complete dictionaries allows a sparse representation because we can decompose the target signal in more than one ways; moreover, the non unique representation of the observed signal gives the possibility of adaptation, the potential of choosing among many representations the one which is most suited to our purpose. Some of the goals are:

Sparsity: we want to obtain the sparsest possible representation of the object, i.e. the one with the fewest significant coefficients. This is why sparse representation helps to capture higher order correlations in the data, it allows for the extraction of a categorically or physically meaningful solution, it gives a parsimonious representation of the signal and, finally, it is easily coupled with dictionary learning techniques.

Discriminative power: the algorithm must produce a good separation of the observed signals $Y = [y_1, \ldots, y_k]$ into classes.

Robustness: small perturbation on y, $\| \Delta y \|$, should not seriously degrade the results; that is, a small perturbation in the observed data should gender a commensurate perturbation in the solution, $\| \Delta x \|$. Robustness is guaranteed by a particular structure imposed to the dictionary called Restricted Isometry Property (RIP) or incoherence (detail information will be given shortly).

Following the standard notation, the matrix format of the general reconstruction or projection step is:

$$y = D \cdot x \tag{2.1}$$

where $D \in \Re^{N \times M}$ is the dictionary, $y \in \Re^N$ the observed signal, and $x \in \Re^M$ is the coefficient vector.

The following picture gives the graphical representation of the problem:

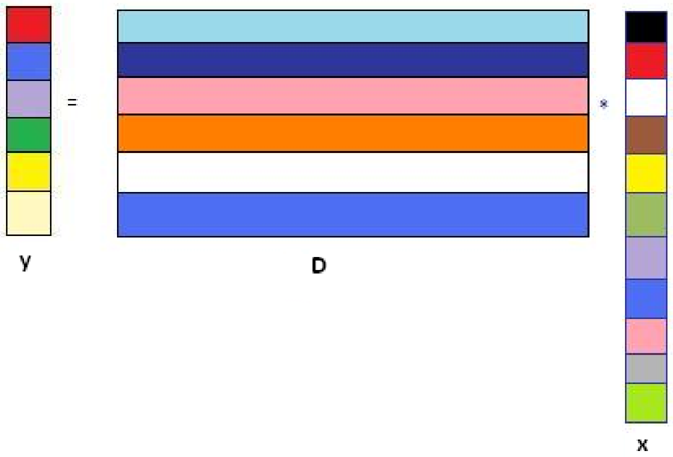


Figure 2.1 The synthesis or analysis problem

When stacking together more than one signal, equation (2.1) becomes:

$$Y = D \cdot X \tag{2.2}$$

where $Y \in \Re^{N \times K}$ and $X \in \Re^{M \times K}$, as depicted by the following picture:

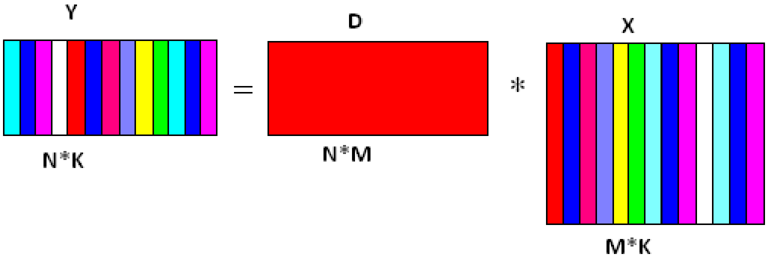


Figure 2.2 The synthesis or analysis problem (matrix wise)

2.1.1 Synthesis versus Analysis

From the synthesis point of view, equation (2.1) has infinitely many solutions at every point; among all possible solutions, we are interested in the one that minimizes the error and has the minimum number of non-zero elements; in other words, we want to minimize the following quantity:

$$\| y - D \cdot x \|_2 + \gamma \| x \|_0 \tag{2.3}$$

where $\| x \|_0$ is the L0 norm, it is simply the count of non-zero elements.

More in details, considering column “j” of matrix *Y*, y_i, during the synthesis step, we need to determine the “best” coefficients along column “j” of *X*, to decompose signal y_j, $\forall j = 1, \cdots, K$, such that:

$$y_j = \sum_{i=1}^{M} d_i x_{i,j} \tag{2.4}$$

or to approximate it using only “t” atoms:

$$y_j = \sum_{i=1}^{t} d_i x_{i,j} + R^t \tag{2.5}$$

where R^t is the residual error obtained by considering only “t” atoms.

On the contrary, from the analysis point of view, we are interested in determining a "good" projection matrix, D, capable of down-sampling the original signal $x \in \Re^M$ into $y \in \Re^N$, $N < M$, without losing critical information.

It is important to notice that synthesis is the operation of building up a signal by superposing atoms, it is the reconstruction step implemented using Basis Pursuit [21] or similar algorithms, while analysis is the projection operator; that is, also the analysis step involves the task of associating coefficients to atoms for any given signal. However, the synthesis and analysis operators are very different because D is a flat matrix, it is not invertible, and the synthesis step cannot use the same coefficients of the analysis step. That is, synthesis and analysis are not self-adjoint operators. Compressive sensing theory is concerned with both the analysis and the synthesis steps, while the Sparse Representation based Classifier (SRC) is strictly related to the synthesis step.

2.1.2 Compressive Sensing Theory

Compressive sensing or compressive sampling (CS) [16], [17], [18] is a novel sensing-sampling paradigm that goes against the common wisdom in data acquisition. CS theory exploits the fact that many natural signals are sparse or compressible in the sense that they have a concise representation when expressed in a proper base; it asserts that it is possible to desire efficient sampling protocols that capture the useful information content embedded in a sparse signal and condense it in a small amount of data.

The actual data acquisition systems, like the one used in the digital camera, follow the Shannon/Nyquist sampling theorem, which tells us that, when uniformly sampling a signal, we must sample at least two times faster than its bandwidth; as a result, we end up with too many samples which are then compressed before being stored or transmitted. More in details, following the Shannon/Nyquist sampling theorem, current data acquisition systems (1) acquire the full M-sample signal x, (2) they compute the complete set of transform coefficients, (3) they locate the s-largest coefficients and discard the $(M-s)$ smallest ones, and (4) they encode the s couples (value, location) of the largest coefficients. As an alternative, the compressive sensing theory proposes a more general data acquisition approach that condenses the signal directly into a compressed representation without going through the intermediate stage of taking all samples. That is, CS theory proposes a new sampling process and asserts that it is possible to recover sparse or compressible signals from far fewer samples than traditional methods use.

Informally, we may state that CS is concerned with s-sparse or s-compressible signals and we may define a signal x as ***sparse*** if all of its entries are zeros except few spikes, and a signal x as ***compressible*** if its sorted coordinates decay rapidly to zero. The following picture gives a graphical representation of sparse and compressible signals in $\Re^3$:

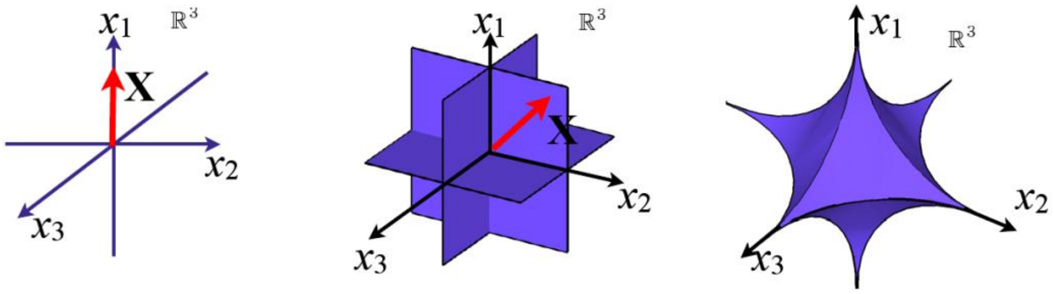


Figure 2.3 Sparse versus compressible signals (courtesy of Boufounos et al. [22])

During the sampling step, to ensure a lossless mapping, it is necessary to impose strong assumptions on the structure of the dictionary D, which must preserve the distance between any two s-sparse vectors; mathematically speaking it must have the ***Restricted Isometry Property*** (RIP) of order $2s$ where s is the level of sparsity of the signal. In formula:

$$(1-\delta_{2s})\| x \|_2^2 \leq \| D_{2s} \cdot x \|_2^2 \leq (1+\delta_{2s})\| x \|_2^2 \qquad (2.6)$$

where δ_{2s} is the restricted isometry constant; in other words, matrix D with the RIP of order $2s$ projects into the down-sampling subspace by preserving information of any s-sparse vector. The RIP property can be verified using the level of incoherence of the dictionary in use: compressive sensing theory states that it is desirable to work with an incoherent matrix, where the ***coherence parameter*** of the dictionary D is defined as follows:

$$DictCoh(D) = \max_{i \neq j} |< d_i, d_j >| \tag{2.7}$$

In other words the coherence is the cosine of the acute angle between the closest pair of atoms. Informally, we may say that a matrix is incoherent if *DictCoh(D)* is small, which is equivalent to say that all atoms of D are very different from each others, they are (nearly) orthogonal vectors.

Moreover, atoms of dictionary D must be incoherent with the basis, $S \in \Re^{M \times M}$, in which $x \in \Re^M$ is assumed to be sparse. That is, in equation (2.1) the original waveform, x, is assumed to be already sparse, when this does not happen, the compressive sensing theory requires the use of a sparsifying matrix, $S \in \Re^{M \times M}$, and the system to be solved becomes:

$$y = D \cdot S \cdot z \text{, where } x = S \cdot z \tag{2.8}$$

obviously $x \in \Re^M$ and $z \in \Re^M$ are the same waveforms from different points of view, but the signal *z* is sparse.

The incoherence of $D \in \Re^{N \times M}$ and $S \in \Re^{M \times M}$ means that the information carried by the few entries of $z \in \Re^M$ is spread all over the *N* entries of $y \in \Re^N$, that is, each sample y_i is likely to contain a piece of information of each significant entry of z_i. These properties allow for a *non adaptive* sampling step, a measurement process which does not depend to the length of $x \in \Re^M$, and produce *democratic* measurements, meaning that every dimension carries the same amount of information, and, therefore, they are equally (un)important.

During the reconstruction step, theoretically, the sparsest solution of system (2.1) can be obtained by solving it as a non-convex constrained optimization method, (2.3), practically this solution is infeasible because it requires an exhaustive enumeration of all $\begin{pmatrix} M \\ s \end{pmatrix}$ possible combinations for the locations of the nonzero entries in x.

Compressive sensing theory showed that, under certain sparsity conditions [23] the convex version of this optimization criterion yields exactly to the same solution. The reason for that is in the geometric shape of the L1 ball, which is represented together with the L2 ball in the following picture:

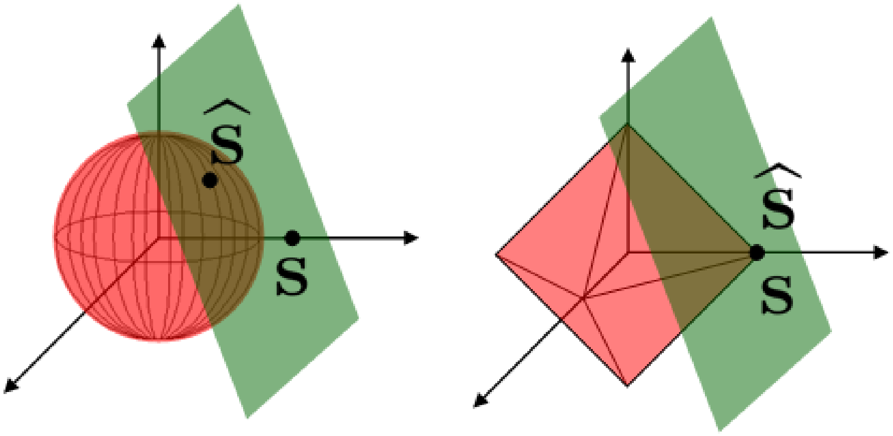


Figure 2.4 L2 versus L1 balls (courtesy of Baraniuk [16])

In Figure 2.4 the green plane is the null space of dictionary D (defined in Paragraph 2.1.3), $\hat{S}$ is the wrong solution recovered by the L2 norm and $\hat{S} = S$ is the correct solution recovered by the L1 norm. That is, knowing that any s-sparse vector lays on an *s*-dimensional hyper-plane close to the coordinate axes, and that we do search for a solution by blowing the interesting balls as far as its border points touch the null space of D, the geometrical shape of the L1 ball, which is pointing along the coordinate axes, allows for perfect recovery of any s-sparse or s-compressible signals, while the L2 ball fails.

Therefore, when the signal to be recovered is sparse enough, compressive sensing theory state that instead of (2.3) we obtain the same solution by solving a much easier convex version:

$$\| y - D \cdot x \|_2 + \gamma \| x \|_1 \tag{2.9}$$

Where $\| x \|_1$ is the L1 norm, equal to the sums of the absolute values of all coefficients.

Interesting to point out that the RIP, or incoherence, property is necessary during the analysis or projection step while during the synthesis step it is the level of sparsity of the wanted solution, which allow for a perfect reconstruction; experimental results proved that the resulting coefficient vector has discriminative power, and, that its sparsity level affects the success rate of the sparse classifier. That is, the idea of using a sparse classifier was born as a possible application of the emerging theory of compressive sensing but, because it is related only to the synthesis step, it does not have the same constraints.

2.1.3 Under vs Over-Determined System of Equations

Let us consider Equation (2.1) $y = D \cdot x$, with $D \in \Re^{N \times M}$, from the linear algebra point of view:

1. If *N=M*, Equation (2.1) is a standard linear system and the solution is unique if D has *full rank*, which is equivalent to say that the *determinant(D)≠0*; the unique solution is the minimum norm solution.
2. If *N<M*, D is a flat matrix having more variables-columns than equations-constraints-rows; the over-complete frame generates an under-determined system of equation. If the equations are inconsistent the system has no solution, otherwise there is an infinitive number of solutions. In the second case, we have a *General Minimum Norm Problem,* that is, among all infinitely many solutions, we are interested in the one minimizing some norms; in case of L2 norm we have the classical *Minimum Norm Problem*, in case of L0 norm, we are looking for the sparsest solution, the one with the minimum number of non-zero coefficients. That is, a flat matrix has a null space, $NS_D = \{v : D \cdot v = 0\}$ of dimension $(M - N)$, and if $y = D \cdot x$ holds also $y = D \cdot (x + v)$ will hold, for all vectors $v \in NS_D$; obviously the true solution is the sparsest one. CS theory proved that, when x is enough sparse, the minimum L0 and the minimum L1 norms return the same solution.
3. If *N>M* matrix D is tall, it has more equations-constraints-rows than variables-columns, and Equation (2.1) has generally no solution. In this case, we are interested in finding x which minimizes the quantity $\| y - D \cdot x \|_2$, and this is the

Standard Least Square Problem, due to the presence of the L2 norm; if, among all possible approximations, we are looking for the one minimizing some norms, then we have a *General Minimum Norm Standard Least Square Problem.* Again, the *Minimum Norm Least Square* problem looks for the coefficient vector x with minimum L2 norm, while we are now interested in the sparsest solution, the one having minimum L0 norm.

Notice that in case of face recognition, N is the resolution of the image while M is the gallery size, the total number of training samples; depending to the experiment under investigation, all cases are possible, but points 2 and 3 are more common, and in both circumstances, we have a general minimum norm problem. Moreover, because columns of D are vectorized faces, and different faces of the same subjects are linearly independent, matrix D has always full rank.

From an algorithmic point of view, when looking for the coefficient vector x having minimum L2 norm, we may use either the pseudo-inverse or the QR factorization of dictionary D; but if we are looking for the coefficient vector x having minimum L1 norm, then we need a *synthesis algorithm*, such as LASSO [24], OMP [25], Augmented Lagrange Multiplier [26], etc. Imposing the minimum L2 norm has a number of computational advantages but it does not provide sparse solutions as it has the tendency to spread the energy among a large number of entries of x instead of concentrate on few coefficients; on the contrary, the minimum L1 norm naturally selects the few and high correlated column of D, and, because of this, the resulting coefficients vector is sparse and has discriminative power.

2.1.4 Synthesis Algorithms

Sparse recovery algorithms reconstruct sparse signals from a small set of non-adaptive and democratic linear measurements. Given $y \in \Re^N$, the measurement vector belonging to a low dimensional space, and the synthesis matrix $D \in \Re^{N \times M}$, $N < M$. The goal is to reconstruct $x \in \Re^M$, the original signal belonging to a high dimensional space, knowing that x is sparse. In other words, the goal is to obtain a sparse decomposition of a signal x with respect to a given dictionary D and the measurements vector y. When we allow for some disturbance elements, we can represent the system with the following formula:

$$\min_x \|x\|_1 \text{ such that } \|y - D\cdot x\|_2 < \varepsilon \qquad (2.10)$$

where ε is a little value. The ideal optimum recovery algorithm needs:

- to be fast: it should be possible to obtain a representation in the order of $\mathrm{O}(N)$ or at most $\mathrm{O}(N\log(N))$ time;
- to require minimum memory storage;
- to provide uniform guarantees: the algorithm must recover all sparse signals, with high probability;
- to be stable: small perturbations of y do not seriously degrade the results.

There are two distinct major approaches to sparse recovery, both of them with pros and cons. The first approach uses greedy methods to compute the support of the signal iteratively. This type of algorithms is usually fast but they do not provide strong guarantees of convergence; examples are Orthogonal Matching Pursuit (OMP) [25] and Matching Pursuit (MP) [27]. The second approach is to cast the sparse recovery problem into a linear programming one; that is, to solve (2.9) instead of (2.3). These methods provide guarantees of convergence and stability, but their running time is not polynomially bounded; examples are Basis Pursuit (BP) [21], and LASSO [24]. A comprehensive review of five representative L1-minimization methods can be found in [26]. Recent developments of the CS theory increased the interest in L1 reconstruction methods and produced new algorithms such as “Compressive Sampling Matching Pursuit” (CoSaMP) in [28], and “A* Orthogonal Matching Pursuit: Best-First Search for Compressed Sensing Signal Recovery” (A*OMP) in [29], with the aim to bridge the gap between the previous two approaches.

2.2 Sparse Representation based Classifier

From a synthesis point of view, the under-determined system (2.1) is used to model different problems where only a small number of information is available out of a very large set of possible sources. Among all possible solutions the sparsest one is preferred because it stores a high description of the data and, therefore, it is useful for the feature extraction stage, or concept generation stage, of pattern recognition problems. Other applications include data compression, high-resolution spectral estimation, direction-of-arrival estimation, speech coding, and function approximation.

The main idea is that many interesting phenomena in nature lie in a smaller, often much smaller, dimensional subspace as compared to the observed signal dimensionality; in other words, the intrinsic magnitude of a signal subspace, where all the variations of the signal occur, is significantly smaller than the ambient dimension. Sparse approximation methods attempt to discover this subspace, and to represent events and objects in that subspace.

Identity recognition is a good case in point: it is assumed that a face can be represented as a sparse linear combination of training samples, which are alternate images of the same subject, and that the resulting combiner coefficients contain discriminative information. Assuming that all training samples of a single class lie on the same subspace, a new down-sampled test image y of class$_i$, y_i, will lie in the linear span of its class's training samples and, therefore, can be represented as a linear combination of only the atoms of D_i, that is:

$$y = D \cdot x_i \tag{2.11}$$

where x_i is a coefficient vectors having all zero entries except for the coefficients of class$_i$, $x_i = [0, \cdots, 0, x_1, \cdots, x_{m_i}, 0, \cdots, 0]^T$; if-when the number of different classes is sufficiently large the representation of y is naturally sparse.

Interesting to notice that the sparse representation based classifier allows for an ad-hoc solution to deal with errors and occlusions. That is, in real world applications the test image, y, could be partially corrupted or occluded, in this case, Equation (2.1) should be modified as:

$$y = D \cdot x + e \tag{2.12}$$

where $e \in \Re^N$ is the error, which is assumed to be sparse, but, the location of non zeros entries is unknown and, obviously, different in every test image. In this case, the simple solution proposed by Wright et al. [14] and [30] is to use an expanded dictionary, $D^{'}=[D \mid I]$; that is, the original dictionary $D \in \Re^{N\times M}$ is expanded to size $\Re^{N\times(N+M)}$ by concatenating a coherent part D of faces to a non-coherent part implemented by the identity matrix, $I \in \Re^{N\times N}$. The use of such an expanded dictionary totally changes the geometry of the system, which was dubbed as cross-and-bouquet (CAB) in [30], due to the fact that the columns of the identity matrix are highly incoherent, they span a cross polytope which captures the error, whereas the columns of D are tightly clustered like a bouquet of flowers. The geometric representation of the expanded dictionary D', built as concatenation of sub-dictionaries, is illustrated in Figure 2.5:

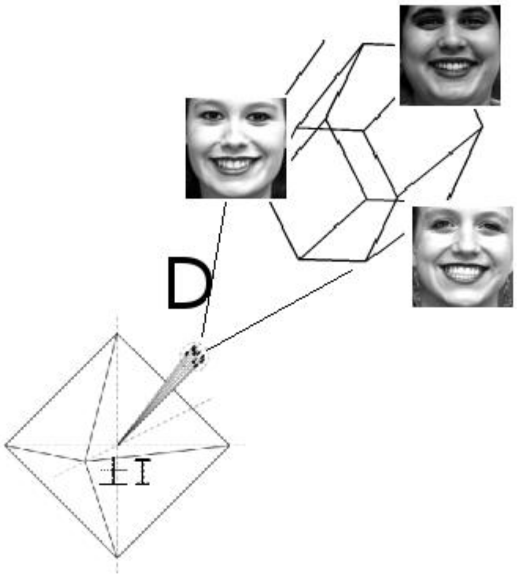


Figure 2.5 The cross-and-bouquet model for face recognition

In other words, the N dimensions of the polytope represent the incoherent noise while the tight bundle captures faces and expressions. This configuration achieves image source separation between face and noise sources.

2.2.1 A case study: Sparse Representation for Face Recognition

The plethora of face recognition methods can be categorized under template-based and geometry-based paradigms. In the template-based paradigm one computes the correlation between a face and one or more model templates for face identity. Methods such as Principal Component Analysis (PCA), Linear Discriminant Analysis, Kernel Methods, and Neural Networks as well as the statistical tools such as Support Vector Machines (SVM) can be put under this category [11], [31]. In the geometry-based paradigm one analyses explicit local facial features and their configurational relationships. The SRC method can be interpreted as a non-linear template matching method.

From a practical point of view, when SRC is applied to face recognition the general framework is as follows: assuming that the database stores m_i faces of subject_i, the column matrix $D_i = [d_{1,} d_2, \cdots, d_{m_i}]$ aligns, column after column, the m_i training samples of class_i; where d_i is a down sampled training face. Having C classes, the synthesis dictionary D must concatenate all training samples of all classes, that is, $D = [D_{1,} D_2, \cdots, D_C] \in \Re^{N \times M}$ with $M = m_1 + \cdots + m_C$, and the problem is modelled as in Equation (2.1), where y is the face to be classified, and x the vector of coefficients to be computed.

It is interesting to point out that, when working with images, dictionary D is *not always flat* because the gallery size, which is the total number of training samples, maybe less than the required resolution. Moreover dictionary D may *not be incoherent*, because it is made up of faces, which are not orthogonal vectors. Nevertheless the performance of SRC is very good, as shown in Chapters 4 and 6.

2.2.2 Algorithms

From an algorithmic point of view the sparse representation based classifier can be implemented in different ways, all variations working on the coefficient vector, x returned by a synthesis algorithm.

The framework is the one introduced in Paragraph 2.1 with C classes, a dictionary D of size $N \times M$, where N is the resolution of the image and $M = m_1 + \cdots + m_C$ is the total number of training samples, and $x = [0, \cdots, 0, x_1, \cdots, x_{m_i}, 0, \cdots, 0]$ is the coefficient vector x restricted to the entries of class$_i$.

During this study, we tried different variations of SRC, some of them presented in the rest of this paragraph.

Distance from Face Space (DFS)

Wright et al. [14] proposed the Distance from Face Space (DFS) variation defined as follows:

$$Class(y) = \min_i \| y - D \cdot x_i \|_2 \qquad (2.13)$$

That is, the class assigned to the test sample y is the one producing minimum residual error, when the face is reconstructed from the class coefficients found by the solution of (2.10). The following table gives the pseudo-code of DFS:

Table 2.1 Pseudo-code of DFS

```
INPUTS: a dictionary $D$, and an observed signal $y$ to be classified

CODE:
1. Normalize all columns of $D$ and the test sample $y$ by imposing unit L2 norm
2. Solve the L1 minimization problem:
       $\min_x \| x \|_1$ such that $\| y - D \cdot x \|_2 < \varepsilon$
3. Compute the residuals, $res_i = \| y - D \cdot x_i \|_2 \ \forall i = 1, \cdots, C$

OUTPUT: Class(y) = the class producing minimum residual
```

Mean of the Class Coefficients (MCC)

The second approach, proposed in this work, is called Mean of the Class Coefficients (MCC) and it is defined simply as:

$$Class(y) = \max_i (E(x_i)) \tag{2.14}$$

Notice that MCC simply calculates the average of the class coefficients and assigns to the test sample y the class having the maximum mean.

The following table gives the pseudo-code of MCC:

Table 2.2 Pseudo-code of MCC

```
INPUTS: a dictionary $D$, and an observed signal $y$ to be classified

CODE:
1. Normalize all columns of $D$ and the test sample $y$ by imposing unit L2 norm
2. Solve the L1 minimization problem:
     $\min_x \| x \|_1$ such that $\| y - D \cdot x \|_2 < \varepsilon$
3. Compute the mean of the class's coefficients, $mean(x_i), \forall i = 1, \cdots, C$

OUTPUT: Class(y) = the class having maximum average
```

Biggest Coefficient DFS

The idea underlying this variation is to consider as decision statistic the biggest coefficient for every class. This is based on the study of the coefficient's distribution (Paragraph 3.2.2) where we point out that the histogram of the biggest coefficient has the least overlap as compared to those of smaller coefficients.

Table 2.3 Pseudo-code of biggest coefficient DFS

```
INPUTS: a dictionary $D$, and an observed signal $y$ to be classified

CODE:

1.  Normalize all columns of $D$ and the test sample $y$ by imposing unit L2 norm

2.  Solve the L1 minimization problem:

3.  $\min_x \| x \|_1$ such that $\| y - D \cdot x \|_2 < \varepsilon$

    Select the biggest coefficient out of every class: $b_i = \max(x_i), \forall i = 1, \cdots, C$

4.  Compute the residuals, $res_i = \| y - D \cdot b_i \|_2, \forall i = 1, \cdots, C$

OUTPUT: Class(y) = the class producing minimum residual
```

Three Biggest Coefficients DFS

This variation is a modification of the previous one, still based on the results of Paragraph 3.2.2; it assumes that the top three coefficients of every class store enough discriminative power and makes classification considering only those coefficients.

Table 2.4 Pseudo-code of three biggest coefficients DFS

INPUTS: a dictionary D, and an observed signal y to be classified

CODE:

```
1. Normalize all columns of $D$ and the test sample $y$ by imposing unit L2 norm
2. Solve the L1 minimization problem:
3. $\min_x \| x \|_1$ such that $\| y - D \cdot x \|_2 < \varepsilon$
   Select the 3 biggest coefficient out of every class: $3b_i = top3(x_i), \forall i = 1, \cdots, C$
4. Compute the residuals, $res_i = \| y - D \cdot 3b_i \|_2, \forall i = 1, \cdots, C$
```

OUTPUT: Class(y) = the class producing minimum residual

Minimum LS Norm

This variation was suggested by [32]; the idea is to solve the usual system $y = D \cdot x$ looking for the minimum L2 norm; that is, instead of using a synthesis algorithm, we calculated the coefficient vector x using either the pinv(D) or the QR factorization of dictionary D; obviously, in both cases, the solution will not be sparse because the minimum L2 norm has the tendency to spread the energy among a large number of entries of x instead of concentrate on few coefficients. We tested the "Minimum LS norm" variation for the emotion classification experiment, but our results were always much worse than the ones presented in Paragraph 6.3.1. Important to point out that also Wright et al. in [14] make a comparison between DFS and the "Minimum LS norm" and show the superior performance of DFS.

Table 2.5 Pseudo-code of minimum LS norm

```
INPUTS: a dictionary $D$, and an observed signal $y$ to be classified

CODE:
1. Normalize all columns of $D$ and the test sample $y$ by imposing unit L2 norm
2. Solve the L2 minimization problem:
       $\min_x \| x \|_2$ such that $\| y - D \cdot x \|_2 < \varepsilon$
3. Compute the residuals, $res_i = \| y - D \cdot x_i \|_2 \; \forall i = 1, \cdots, C$

OUTPUT: Class(y) = the class producing minimum residual
```

Different Cardinality Class DFS1

This variation was suggested by the idea that the performance of DFS can be affected by the different cardinalities of the classes. This version of DFS considers only the top "min_card" coefficients of every class, where "min_card" is the minimum of the number of representatives among all the C classes. In pseudo-code:

Table 2.6 Pseudo-code of different cardinality class DFS1

INPUTS: a dictionary D divided into classes of cardinality (card$_1$,…, card$_C$), and an observed signal y to be classified

CODE:

```
1. Normalize all columns of $D$ and the test sample $y$ by imposing unit L2 norm
2. Solve the L1 minimization problem:
       $\min_x \| x \|_1$ such that $\| y - D \cdot x \|_2 < \varepsilon$
3. min_card=min(card$_1$,…, card$_C$), $t_i$ = top “min_card” coefficients of $x_i$, $\forall i: 1, \cdots, C$
4. Compute the residuals: $res_i = \| y - D \cdot t_i \|_2 \ \forall i = 1, \cdots, C$
```

OUTPUT: Class(y) = the class producing minimum residual

Different Cardinality Class DFS2

This variation was created with the aim to increase the robustness of DFS in case of classes with different cardinality. It multiplies the DFS score by a normalization coefficient, in pseudo-code:

Table 2.7 Pseudo-code of different cardinality class DFS2

INPUTS: a dictionary D divided into classes of cardinality ($card_1$,…, $card_C$), and an observed signal y to be classified

CODE:

```
1. Normalize all columns of $D$ and the test sample $y$ by imposing unit L2 norm
2. Solve the L1 minimization problem:
   $\min_x \|x\|_1$ such that $\|y - D \cdot x\|_2 < \varepsilon$
3. Compute the residuals: $res_i = \|y - D \cdot x_i\|_2 \ \forall i = 1, \cdots, C$
4. norm_coeff=(dim(x)-$card_i$)/ dim(x); $score_i$= norm_coeff × $res_i$, $\forall i : 1, \cdots, C$
```

OUTPUT: Class(y) = the class producing minimum score

Nearest subspace SRC

In this variation the DFS algorithm is applied separately to every subspace. In case of the emotion classification challenge every test subject will have 7 dictionaries, one per emotion, with different cardinality depending to the total number and the type of emotional faces of the current test subject. Instead of putting all these dictionaries together, D=[AnDict, CoDict, DiDict, FeDict, HaDict, SaDict, SuDict], and solve the system $y = D \cdot x$, we do now solve 7 separate systems.

Obviously, all coefficient vectors are NOT sparse any more neither with a synthesis algorithm, because, at any time, the dictionary stores only one class; classification is done as usual by assigning the test sample to the nearby class.

- *Variation1*: it solves the 7 systems separately by imposing minimum L1 norm.

- *Variation2*: it solves the 7 systems separately by imposing minimum L2 norm.

Both variations were tested in the emotion classification experiments; in both cases the obtained performance is worse than the one documented in Chapter 6.

MCC with Absolute Value

The MCC variation previously introduced has a very good performance but it lacks from mathematical background; by introducing the absolute value, the algorithm calculates the mean of the L1 norm of the class's coefficients, but this variation is not as successful as the original MCC.

Table 2.8 Pseudo-code of MCC with absolute value

```
INPUTS: a dictionary $D$, and an observed signal $y$ to be classified

CODE:
1. Normalize all columns of $D$ and the test sample $y$ by imposing unit L2 norm
2. Solve the L1 minimization problem:
   $\min_x \| x \|_1$ such that $\| y - D \cdot x \|_2 < \varepsilon$
3. Computer the mean of the class's coefficients, $mean(abs(x_i)), \forall i = 1, \cdots, C$

OUTPUT: Class(y) = the class having maximum average
```

The most successful variations are DFS and MCC, their performance is generally the same; for this reason, in the sequel, we will refer to them in the tables simply as SRC.

CHAPTER 3

STUDY OF THE CHARACTERISTICS OF SRC

In this chapter we present some parameters, which are considered critical for the performance of the sparse representation based classifier, and we introduce other new critical parameters. In Chapter 4, for some relevant experiments, we will monitor those measurements with the aim to determine “if” and “how” they are correlated to the success rate of SRC. Furthermore, this chapter runs experiments to investigate some statistical properties of the coefficients’ vector. All experiments are based on a particular selection done on the cropped sub-directory [33] of the Extended Yale B database [34]: 38 subjects, 59 images per class divided into 30 training and 29 test samples, which results in 30×38=1140 training and 29×38=1102 test samples; working at low dimension 504 the dictionary’s size is 504×1140. The Extended Yale B database has illumination effects, but in this case, the careful division into training and test sets ensures that the training set sees all elevation and azimuth angles; that is, it is an informed experiment with high success rate.

3.1 Critical Parameters

3.1.1 Sparsity Level

The sparsiy level (SL) of the coefficient vector is normally considered as one of the main factors reflecting the performance of the classification algorithm. In order to monitor the correlation between sparsity and the performance of SRC it is first of all necessary to define a formula to calculate the level of sparsity of a given vector. The first straightforward way is the number of non zero coefficients, which bring to the definition of sparsity level (SL):

$$SL = \frac{number\_non\_zero\_coeff}{gs} \tag{3.1}$$

Where "gs" is the gallery size, which is the total number of training sample equal to the "enrolment size × number of classes". The range of SL= (0, 1], where "near to 0" means sparse.

In this thesis we conjecture that the success rate of SRC is correlated to the level of compressibility of the coefficient vector and not to its level of sparsity, and because the SL formula can measure only the level of sparsity of a vector, its outputs are not correlated to the performance of SRC. To better explain this concept, let us consider two sets of vectors in $\Re^2$: set A={a_1, a_2, a_3}={(0,1), (1,0), (0.9,0.1)} stores two sparse and one compressible signal, while set B={b_1}={(1,1)} has a non-sparse vector. Because SL(a_3)=SL(b_1), the SL formula does not distinguish between the compressible and the non-sparse vector and returns the same value in both cases.

On the contrary the following formula measures the Compressibility Level (CL) of the input vector:

$$CL = \frac{\sqrt{gs} - \left(\frac{\sum_j |x_j|}{\sqrt{\sum_j x_j^2}} \right)}{\sqrt{gs} - 1} \tag{3.2}$$

The range of CL= [0, 1], where "1" means compressible. In order to clarify the meaning of CL formula, let us consider the two extreme cases and the example given before:

Case 1: The vector has only one coefficient different from "0".

This is the trivial case where $\left(\frac{\sum_j |x_j|}{\sqrt{\sum_j x_j^2}} \right) = 1$, which produces CL=1

Case 2: The vector has all coefficients equal to "1". In this case

$\left(\frac{\sum_j |x_j|}{\sqrt{\sum_j x_j^2}} \right) = \frac{gs}{\sqrt{gs}} = \sqrt{gs}$ which produces CL=0.

Case 3: Because L1(a_1)=L1(a_2)=L1(a_3) the sparse and compressible vectors will have same CL value.

It is important to underline that the CL formula is theoretically correct but not sensitive enough because the synthesis algorithm always returns a sparse solution, which does not contain all "1".

In this thesis we conjecture that SL is not a critical parameter of SRC while CL is correlated to the success rate of the classifier; we will test this hypothesis in Chapter 4.

3.1.2 Coherence of the Dictionary

Compressive sensing theory rather to work with an incoherent dictionary $D=[d_1, \cdots, d_i, \cdots, d_j, \cdots, d_M]$, where the coherence parameter *DictCoh* was defined in (2.7): the coherence is the cosine of the acute angle between the closest pair of atoms, the closes columns of the dictionary. Informally, we may say that a dictionary is incoherent if *DictCoh* is small, which is equivalent to say that all atoms of dictionary, D, are very different from each other, they are nearly orthogonal vectors. The range of *DictCoh* is [0,1], where "0" means incoherence.

In this study, we conjecture that the coherence of the dictionary is not a critical parameter of SRC; we will test this hypothesis in Chapter 4.

3.1.3 Mutual Coherence Dictionary-Test Sample

In this work we introduce a new parameter to monitor the mututal coeherence between the dictionary and the test sample, we call it *DictTestCoh*. For every test sample, we calculate its mutual coherence with the dictionary defined as:

$$DictTestCoh(D, test_j) = \min |<d_i, test_j>| \tag{3.3}$$

$\forall\, i = 1, \cdots, M$ and $\forall\, j = 1, \cdots, K.$ In this thesis we conjecture that nearby test samples have higher recognition rate; we will test this hypothesis in Chapter 4.

3.1.4 Confidence Level

The confidence level is the second critical parameter introduced in this study, it measures the amount of conviction of the classifier as the distance between the score of the winner class with the one of the runner up. To better illustrate this concept, we run the pre-definite experiment on the Extended Yale B database (38 classes) and we saved in Figure 3.1 the set of residuals which allow for correct classification of a test sample and in Figure 3.2 the set of residuals which bring to misclassify the current test sample.

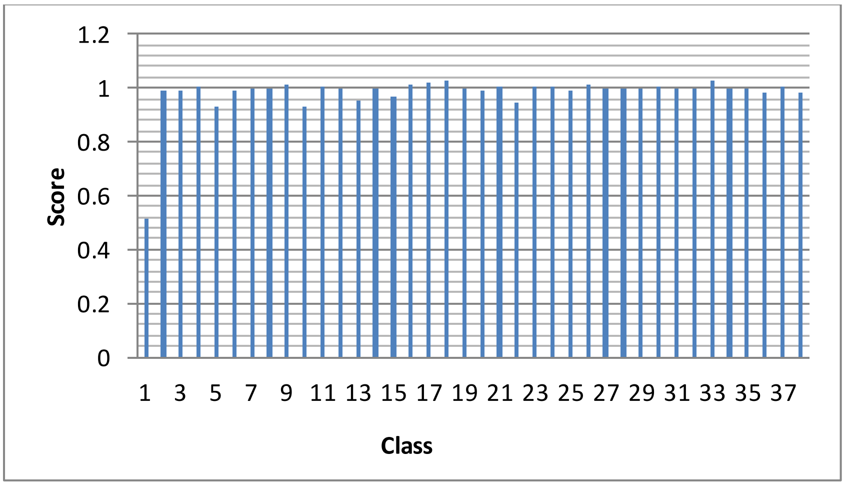


Figure 3.1 Residuals of a correct classified test sample

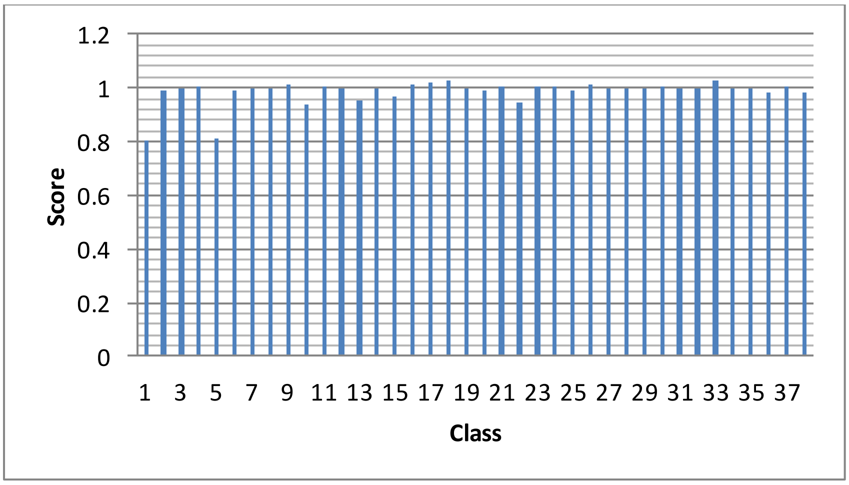


Figure 3.2 Residuals of a misclassified test sample

While in Figure 3.1 the score of class 1 is obviously the little one and there is no uncertainty in declaring class 1 as the winner class, in Figure 3.2 the decision of SRC is not as clear as before because the score of the winner and the one of the runner up are near to each other, their difference is minimal, and the classifier is not as confidence as in the first case.

In this thesis we conjecture that correct classified test samples have bigger confidence level than misclassified one, where the level of confidence is measure with the following formula:

$$DeltaScore = Score(winner) - score(runner\_up) \tag{3.4}$$

We will test this hypothesis in Chapter 4.

3.2 Statistical Studies of SRC

3.2.1 Coefficients Correlation Matrix

The aim of this study is to investigate the intra class correlation of the coefficients of the target class. In order to do that, we run the pre-defined experiment on the Extended Yale B database (38 classes with enrolment size 30 and 29 test samples per class) while saving, for every test sample, the set of all target coefficients; this result in a coefficient matrix of size (38*29) × 30= 1102 × 30, where columns are the fixed 30 dimensions, while rows are the number of samples, 1102 in this experiment. We created the correlation matrix of non ordered coefficients, and we compared it with the correlation matrix of algebraic ordered coefficients and the one of coefficients ordered by their absolute values. Before plotting all correlation matrices were normalized by imposing “1” to the diagonal elements:

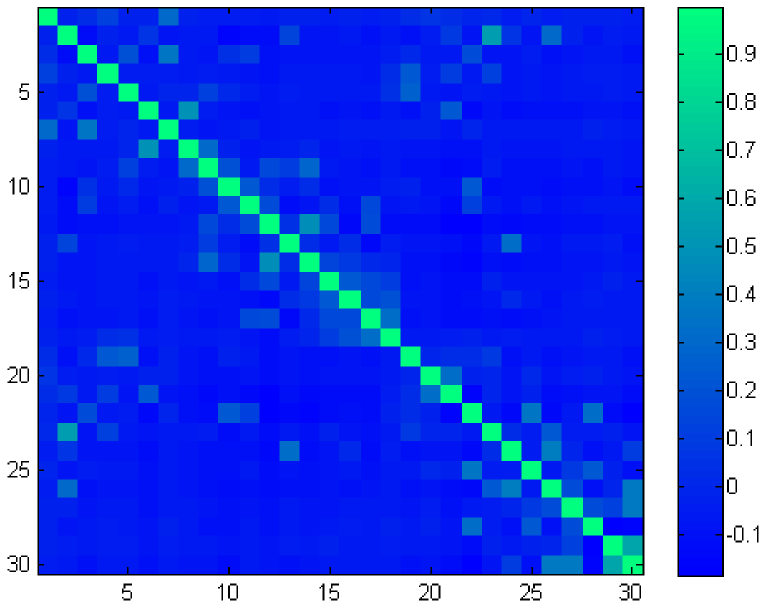


Figure 3.3 Correlation matrix of the 30 non-ordered coefficients of the target class

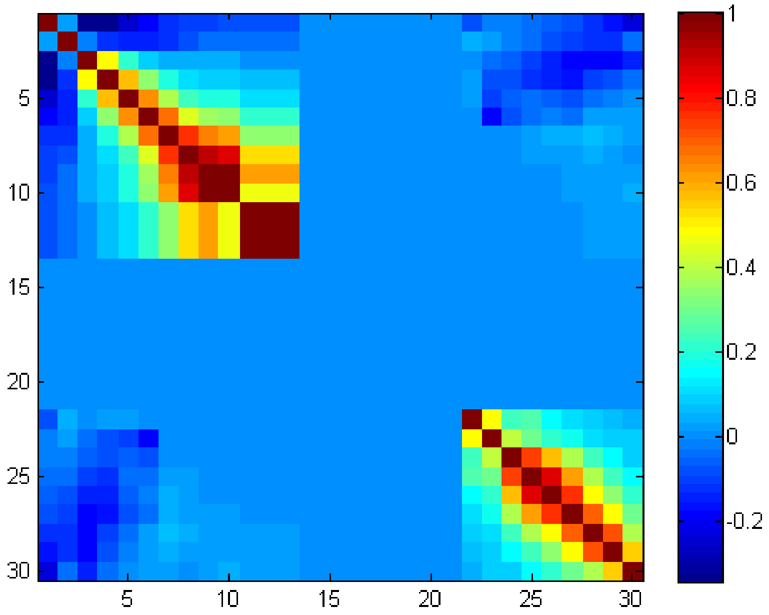


Figure 3.4 Correlation matrix of the 30 coefficients of the target class ordered by their algebraic values

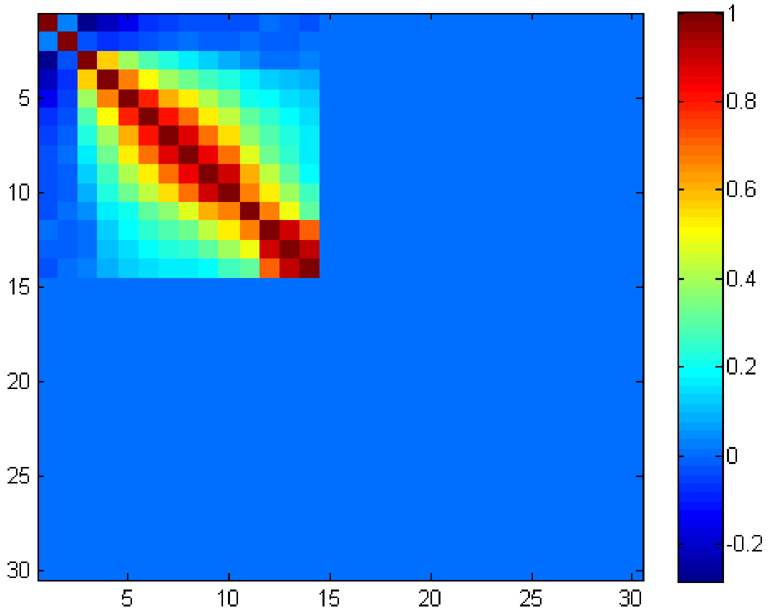


Figure 3.5 Correlation matrix of the 30 coefficients of the target class ordered by their absolute values

Looking at the previous pictures we can see that:

- The majority of values in Figure 3.3 are nearby zeros, which means that the coefficients of the target class are originally uncorrelated;
- The central rows and columns of Figure 3.4 are all “0”, and this is justified by the high level of sparsity of the target coefficients. That is, in case of algebraic ordered coefficients, the values of the central dimensions, [14, 21], are all “0”, because the L1 norm returned a sparse solution, where most of the coefficients are “0”, and, when ordered in an algebraic way, logically, the central dimensions will be “0”;
- Figure 3.5 shows the correlation matrix of the target coefficients ordered by their absolute values: only the first dimensions are different from “0.

This study shows that correlation matrices are heavily affected by the high level of sparsity of the coefficient vectors.

3.2.2 Coefficients Histogram

The aim of this experiment is to investigate the discriminative power of the coefficient vector. In order to do that, we plotted the coefficient’s distribution of the target class against the one of the impostor classes; as usual we worked with the pre-definite experiment briefly introduced in the previous paragraph: Extended Yale B database, 38 classes with enrolment size 30 and 29 test samples per class.

After having find out the values of the 30 targets coefficients and the ones of the (30×37) impostors' coefficients, we chose the biggest range for both target and impostor bars, [-0.5:0.1:1], and we used these 16 bins to plot the coefficient distribution: for every test sample to be classified the synthesis algorithm returns a set of coefficients, 1140, logically divided into 38 classes, 30 coefficients per class. Knowing the true class of the test sample, we considered first the 30 coefficients of the target class, we ordered them and we updated the *TargetSizeBar*: the first stronger coefficient contributes with one vote to the correct bin of the first column, .., the weakest coefficients contributes with one vote to the correct bin of the last column. In a similar way, the 37 impostor classes of every test sample were processed in sequel to fill in the *ImpostorSizeBar*. The following pictures show the histograms of relevant coefficients; in all pictures (x,y) coordinates are (Coefficient Distribution, Size Bar):

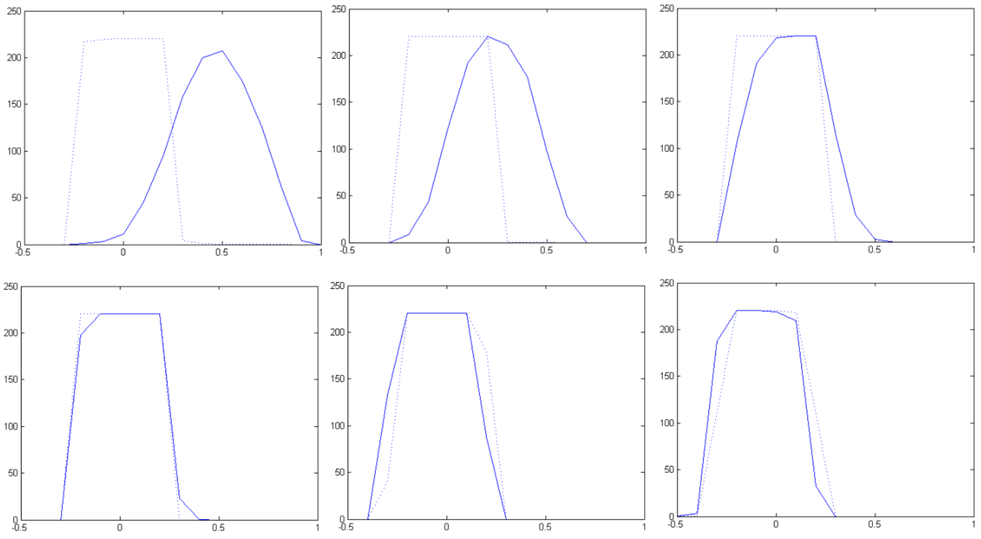


Figure 3.6 Histograms of target coefficients no.1, 2, 3, 4, 29, and 30

Interesting coefficients are until no.4 and after 29, but still there is big overlapping. The biggest coefficient looks to be the one storing the highest discriminative power but the SRC variation using only that coefficient, the "Biggest Coefficient DFS", is not as successful as DFS or MCC variations; this is probably due to the different basis elements selected.

CHAPTER 4

SRC WITH CORRUPTED FACES

The aim of this study is to explore the capabilities of the sparse representation based classifier. In this chapter we address the problem of 2D face classification under adverse conditions; the study is conducted experimentally using the cropped sub-directory [33] of the Extended Yale Face B database [34], which stores cropped and aligned faces with illumination changes. This dataset allows for investigating the performance of SRC in case of disturbance elements such as low resolution, variable gallery size, noise, changes in illumination and geometric distortions. SRC results are compared against the ones of the Fisher classifier, which is used as benchmark [35]. While running the experiments we monitored also the parameters introduces in Chapter 3, so as to detect "if" and "how" they are correlated to the success rate of SRC. For every trial, we reported its success rate together with its memory requirement, as a measure of the complexity of the experiment. The obtained results testify the good performance of SRC, due to its inner robustness derived from the presence of the L1 norm, which is beneficial in mitigating various adverse effects.

4.1 Classification of Corrupted Faces via SRC

Automatic face recognition is still an open issue in the computer vision community, due to the presence of many disturbance elements such as low resolution, illumination, expression, pose, and occlusion. In this section we concentrate on (1) the choice of down-sampling subspaces and the effect of low resolution on the classifiers' performance, (2) we study the effects of the degree of over-completeness, (3) and illumination changes, (4) we explore the robustness of SRC to shifts, in-plane rotations and scaling, and (5) we evaluate the robustness of SRC algorithm to noise. In case of shifts and rotations, we propose a simple trick to increase the inner robustness of SRC by using a shifted and rotated dictionary. In most of the experiments the performance of SRC is compared with the one of the Fisher classifier, with is used as benchmark.

4.2 Description of the Database

We used the cropped face sub-directory [33] of the Extended Yale Face B database, [34] to test the robustness of SRC-based face recognition in adverse conditions. The cropped sub-directory of the database consists of aligned and cropped face images of 192*168=32,256 pixels. For each of the 38 subjects, there is a subdirectory consisting of between 59 to 64 images of that person under various illumination changes. These images differ in azimuth and elevation angles of illumination, where these angles are spaced by a few tens of degrees. In total we worked with 2432 face images. Figure 4.1 shows how to measure elevation and azimuth angles:

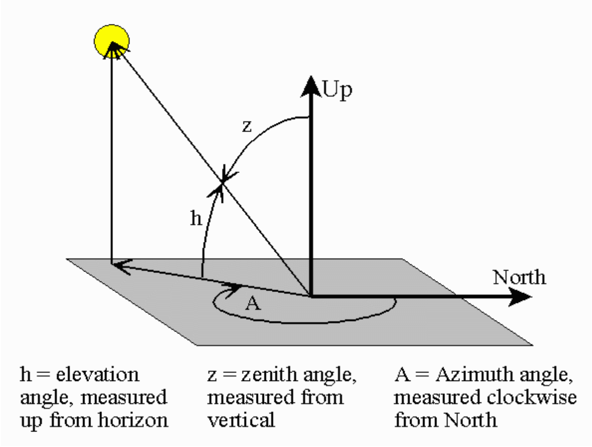

Up
z
h
North
A
h = elevation angle, measured up from horizon
z = zenith angle, measured from vertical
A = Azimuth angle, measured clockwise from North

Figure 4.1 Elevation and azimuth angles

4.3 Experiments

The cropped subdirectory of the Extended Yale B database stores cropped and aligned faces with changes in illumination. This database is suitable to study the robustness of SRC to illumination changes as well as it allows for the insertion of a controlled amount of noise and in-plane rotation. Figure 4.2 stores samples of faces used in our experiments:

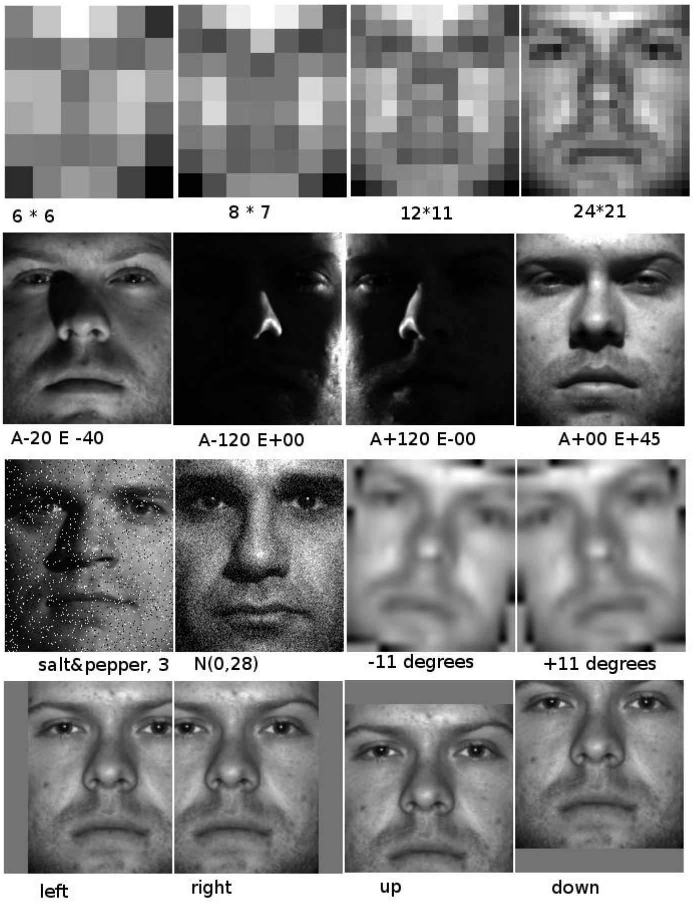


Figure 4.2 Samples of faces under adverse conditions

More in detail, in the 1st row of Figure 4.2 there are images at different resolutions, the 2nd row stores faces under various illumination effects, the 3th row shows a 3% salt&pepper noisy image, a Gaussian noisy face, N(0,28), and two rotated pictures, and the 4th row stores 3 pixels shifted faces in all directions.

Unless otherwise stated, in all experiments we selected 30 training and 29 test images per class, and we worked with decimated images of size 504 (corresponding to images of size 24×21 after down-sampling by a factor of 8 in each direction). Thus the resulting dictionary D has size 504 × 1140, since there are 30 images for each of the 38 subjects (classes). It follows that the Fisher classifier, which is our benchmark, has 37 discriminant planes.

4.3.1 Down-Sampling Subspaces

While the original Extended Yale B face images are 192×168=32.256-dimensional, we investigated the extent the dimensionality could be lowered without compromising the performance; we run experiments at low dimensions 36, 56, 132 and 504 corresponding the a shrink factor of 32, 24, 16 and 8. Among dimensionality reduction methods, we considered decimation, random projections and PCA [36] subspace representations. Decimation is simply the operation of low-pass filtering and sub-sampling. Random projections use random, unit-variance, zero-mean Gaussian vectors; due to the presence of the random factor, the given performance is the median value of 5 trails. The rationale of representing faces by their random projection is compressed sensing theory; accordingly it was shown that signals that are intrinsically low-dimensional can be reconstructed using constrained sparse optimization from far fewer random projections as compared to their Nyquist rate. PCA subspace representation is obtained by projecting the faces onto PCA basis vectors and reconstructing them with the fewer most energetic ones. Thus columns 36, 56 .. 504 in Table 4.1 indicate that faces were classified with 36, 56 .. 504 PCA bases.

For what is concerning memory requirements, the 1st dimension of the dictionary changes with the image resolution, from 36 to 504, while the 2nd dimenstion is fixed to (30 × 38), which is the gallery size of every experiment; at every trial the number of test samples is (29 × 38).

Table 4.1 Effects of down-sampling subspaces and image resolution on classifiers' performance (%)

	SRC(36)	SRC(56)	SRC(132)	SRC(504)	Fisher (37)
Decimation	94.61	97.10	98.92	99.50	95.28
Random Projection	94.90	97.31	98.90	99.34	94
PCA	95.64	97.37	97.82	97.82	95.58

The results in Table 4.1 need some interpretation. First, the Sparse Representation based Classifier (SRC) is not affected by the choice of the down-sampling subspace; in the sequel we will use the decimation subspace, which does not require to run multi-trials. Second, it is surprising to see that keeping all training samples in the dictionary and doing classification by their linear combiner coefficients is much better as compared to amassing all the training data information in a statistical model, that is, class means and variances. In fact, with faces decimated to size 24×21 the SRC method achieves 99.50% recognition rate, 4 percentage points above that of FDA. The price to pay for this higher performance is the need to store and operate on all the sample feature vectors. These results should be interpreted with some precaution though: the Ext. Yale B database provides a dense sampling of the face manifold under illumination directions so that any test face can find a close companion image, in other words, for each test image, there are training faces that differ slightly in azimuth or elevation angle of the illumination direction.

4.3.2 Effect of the Degree of Over-completeness

In this set of experiments we considered the effect of the dictionary size on the performance of the SRC classifier and concomitantly of the training data size for the Fisher classifier. As we increase the number of sample images per subject we have a richer training set. To this effect, we changed the gallery size in steps of 5, from 5 to 50, using each time random selection of subsets of the gallery. Thus for example, at one extreme we selected 5 faces for training and 55 for testing; at the other extreme 50 faces for training and 10 for testing. In this experiment, we used only 24 × 21 (504 pixels) decimated face images, and the size of the dictionary changes from 504 × (5 × 38) to 504 × (50 × 38). The training hence testing subsets are randomly selected from the given YaleB subdirectory; that is, every trial is based on a random permutation, which partitions the images of every subject into training and test sets. In order to limit the effect of the random choice, the given performance is the median value over 5 trials. Table 4.2 stores the obtained performance (%):

Table 4.2 Effects of the gallery size on classifiers' performance

Enrol. size	5	10	15	20	25	30	35	40	45	50
SRC	78.7	86.3	94.6	96.4	98.4	99.2	99.1	99.5	99.4	99.7
Fisher	69.9	75.5	81.9	90.6	95.3	95.5	97.7	97.8	98.9	99.4

Figure 4.3 plots the results of Table 4.2:

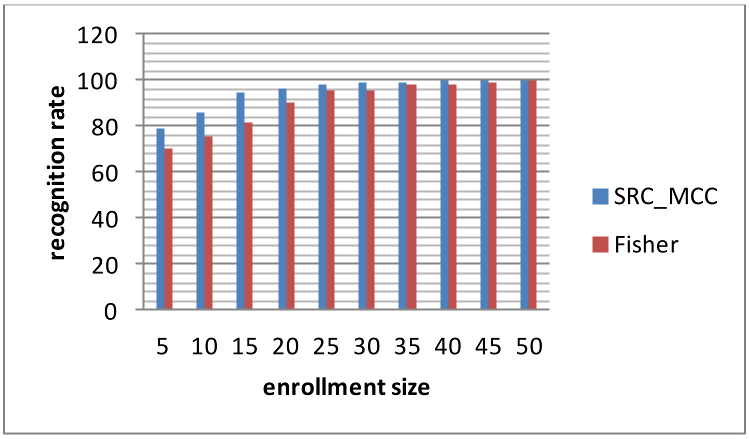


Figure 4.3 Effects of the gallery size on classifiers' performance (plot of Table 4.2)

These results show that both algorithms, SRC and Fisher, do increase their recognition rates as the enrolment size increases. It is surprising to notice the superior performance of a data driven method: much more robust for low size enrolment and still better than Fisher also in case of large degree of over-completeness. That is, the recognition rate of SRC is 9 percentage points above Fisher in case of enrolment size 5, and still slightly better than LDA with 50 training pictures per subject.

4.3.3 Noise

We evaluated the robustness of the SRC algorithm to noise that simulates impairments due sensor noise. We run a blind experiment; we added Gaussian and salt&pepper noises to the test images, already down-sampled by a factor of 8. Gaussian noise is gauged according to Peak Signal to Noise Ratio (PSNR) and salt&pepper noise is characterized by the percentage of pixels contaminated. For what is concerning memory requirement, in all trials the size of the dictionary is 504 × (30 × 38) and the number of test samples is 29 × 38. The results stored in Table 4.3 show the superior performance of SRC also in case of noise. As expected, Fisher, which is a discriminative method is not robust to noise. The impressive result is that SRC performs always better than Fisher and it is also robust to noise; that is, with a PSNR of 20 the performance of SRC is only 3 points less than the original recognition rate. Moreover, the initial gap with Fisher increases from 3 percentage points up to 76. Figure 4.4 and 4.5 plot the obtained performance (%) versus PSNR and salt&pepper percentage.

Table 4.3 Performance (%) under Gaussian (left) and salt&pepper noise (right)

PSNR	SRC (504)	Fisher (37)	**salt&pepper (%)**	SRC (504)	Fisher (37)
Inf	98.28	95.92	0	98.28	95.92
47.81	98.28	95.46	1	97.19	54.36
38.91	98	92.47	3	96.19	24.89
31.58	97.55	82.30	7	93.74	11.80
25.67	97.37	51.18	10	91.65	10.07
19.39	96.1	20.15	20	80.40	5.54

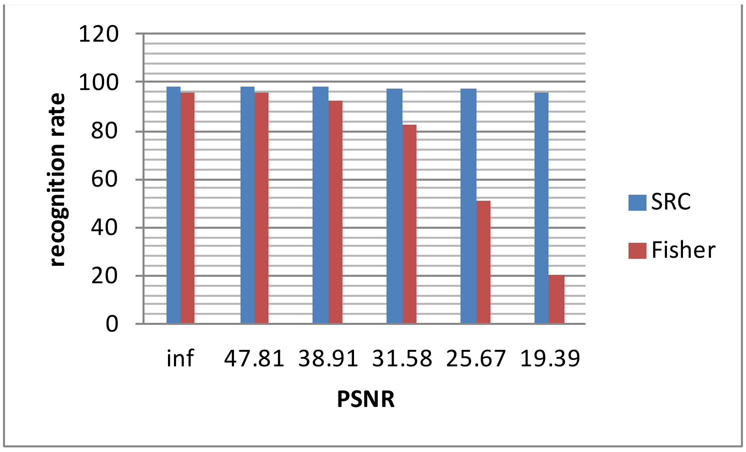


Figure 4.4 Performance (%) under Gaussian noise (plot of Table 4.2, left side)

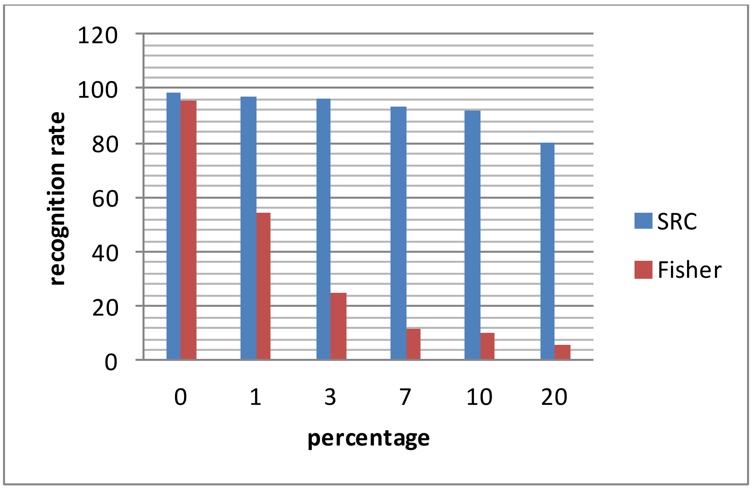


Figure 4.5 Performance (%) under salt&pepper noise (plot Table 4.2, right side)

4.3.4 Illumination Compensation Capabilities

We investigated the robustness of the classifiers against illumination effects and whether it was possible for the detector to operate with faces which are subjected to unseen illumination effects. In order not to bias the results we did not apply any illumination normalization algorithm, and we created set of cardinality 19, which is the minimum number of available faces per subject with negative elevation angle. For this purpose we carried out the following two experiments:

1. Azimuth Angle Segmentation: We grouped the Ext. Yale B database images into two sets, which consisted respectively of all images with positive azimuth, A^+=[+25, +130], and all images with negative azimuth, A^-=[-20, -110], independent of their tilt (elevation) angles.

2. Elevation Angle Segmentation: We grouped the Ext. Yale B database images into two sets, which consisted respectively of all images with positive elevation angles, E^+=[+00, +90], and all images with negative elevation angles, E^-=[-10, -40], independent to their azimuth.

In both cases we exchange train and test sets as illustrated in the following table:

Table 4.4 Training and test sets of experiments on illumination

Azimuth-Elevation Angle Range	Exp1	Exp2	Exp3	Exp4
A^+=[+25, +130]	train	test		
A^-=[-20, -110]	test	train		
E^+=[+00, +90]			train	test
E^-=[-10, -40]			test	train

For what is concerning memory requirements, in all experiments the size of the dictionary is 504 × (19 × 38) and the number of test samples is 19 × 38. Table 4.5 stores the performance of SRC and Fisher classifiers together with the behaviour of the critical parameters introduced in Chapter 3: columns “SL”, and “CL” show the mean values of sparsity and compressibility levels, “DictCoh” is the coherence of the dictionary, “DictTestCoh” is the mutual coherence between the dictionary and the test matrix, and “DeltaScore” measures the level of confidence of the classifier.

Table 4.5 Illumination compensation capability

	SRC (%)	Fisher (%)	SL (mean)	CL (mean*100)	DictCoh (*100)	DictTestCoh (mean*100)	DeltaScore (mean*100)
Exp1	33	24	49.45	88.25	98.78	83.31	14.97
Exp2	25	24	43.5	87.7	99.69	64	7.85
Exp3	93	90	52.64	92.60	99.69	95.04	56.00
Exp4	81	77	53.49	90.28	99.73	91.79	48.31

The results of Table 4.5 show that SRC is more robust than Fisher classifier to changes in both azimuth and elevation angles. It is important to notice that the better performance of SRC for the experiments on elevation angle segmentation, Exp3 and Exp4, is simply due to the particular structure of the database which varies the azimuth angle in a wide range, from [-130, + 130], while keeps most of the picture into the elevation angle's range [-45, +45]. That is, the azimuth angle experiment is more challenging than the elevation angle one.

As expected the sparsity level of the coefficient vector, x, has an irregular behaviour which does not reflect the successful rate of the classifier. On the contrary, the compressibility level decreases together with the recognition rate. In every experiment the dictionary changes but its level of incoherence is always very high; that is, *DictCoh* is always nearby “1”, which means that the dictionary is highly coherent, as expected because it is made up of faces. These results show that, the success rate of the dictionary does not depend to its inner characteristics, rather than it is related to the correlation level between training and test data: when training and test samples are nearby each other, the *DictTestCoh* is “0.95” and the performance is 93%, when train and test samples are different the mean value of *DictTestCoh* drops to “0.64” together with the performance of SRC, 25%. Finally, *DeltaScore* appears to be the most sensitive parameter: starting from “0.56” in case of experiment 3 with a success rate of 93% and dropping to “0.07” in case of experiment 2 with a success rate of 25%.

4.3.5 Geometric Distortion: shift, in-plane rotation and scaling

In real applications images rarely are in perfect registration, that is, presented fully frontally and in the correct scale and position. In order to test the robustness of the classifiers against mis-registration effects, we perturbed the faces with shifts, in-plane rotations and zoom. To preclude the confounding effects of illumination, we previously selected faces with nearly frontal illumination, that is, those having azimuth in the range of [-25, +25]. This results in 23 pictures per class, which are then randomly divided into training (20 images) and test (3 pictures) sets. All experiments are repeated 5 times and the given recognition rate is the median value.

In the shift experiment we worked with original sized test images so as to consider also fractional shifts. The 192 × 168 test samples were shifted from 2 to 16 pixels in all directions (up, down, left, right); as usual classification was then performed in the low dimensional space, 504. In the rotation experiment down-sampled test faces were rotated in-plane by ±1, ±3, ±5, ±7, ±9, and ±11 degrees. In the zoom experiment, down-sampled test images were zoomed by a scale factor in the range [0.7:0.1:1.3]. To avoid imaging artefacts, image parts overawing the 24×21 were cropped; conversely, if background was disclosed it was padded with average image gray level.

For what is concerning memory requirements, in all experiments the size of the dictionary is 504 × (20 × 38), while the number of test samples changes in case of shift, (3 × 4 × 38), rotation, (3 × 2 × 38), and zoom, (3 × 38).

The performance (%) of shifted, rotated and zoomed images is stored in the following table and plotted into Figures 4.6, 4.7 and 4.8:

Table 4.6 Blind experiments on shift (left), rotation (centre), zoom (right)

Shift			**Rotation**			**Zoom**		
pixel	SRC	Fisher	Degrees	SRC	Fisher	scale	SRC	Fisher
2	100	100	±1	100	100	0.7	3.51	4.39
4	100	99.56	±3	100	100	0.8	14.91	6.23
6	96.93	78.51	±5	99.12	73.25	0.9	72.81	7.89
8	75.66	35.31	±7	85.09	34.21	1	100	100
10	50.66	13.82	±9	64.91	15.35	1.1	47.37	13.16
12	29.39	6.8	±11	45.18	7.46	1.2	26.32	6.14
14	18.86	3.95				1.3	5.26	4.39
16	13.16	3.73						

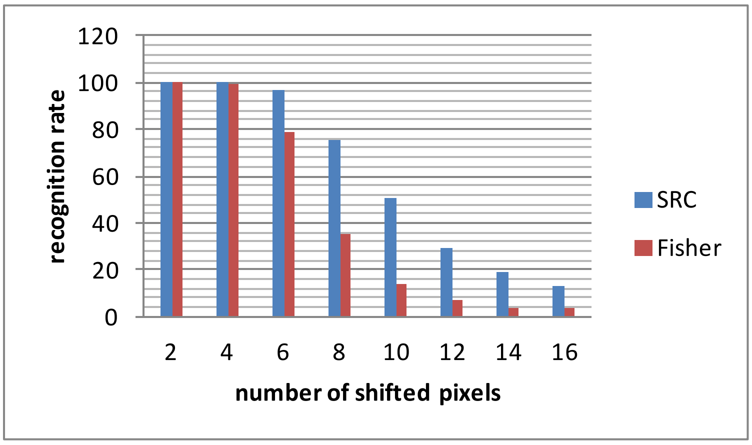


Figure 4.6 Effects of the shift geometric distortion (plot of Table 4.6, left side)

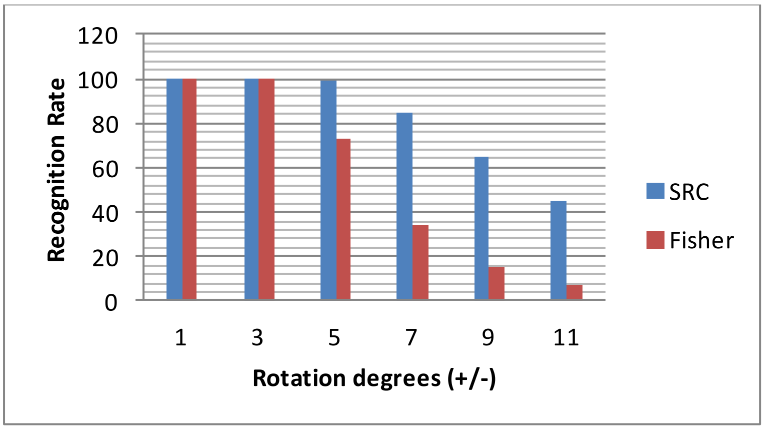


Figure 4.7 Effects of the rotation geometric distortion (plot of Table 4.6, center)

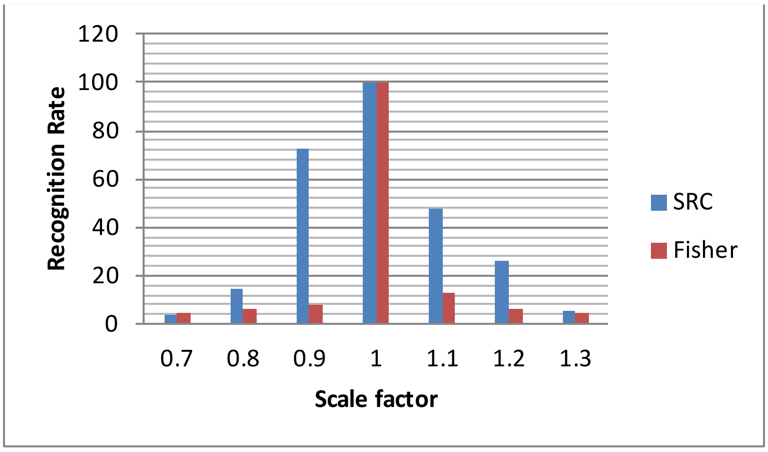


Figure 4.8 Effects of the zoom geometric distortion (plot of Table 4.6, right side)

These results show that both algorithms are heavily affected by geometric distortions. For SRC the problem is tackled in [9] were Wagner et al. present a "deformable SRC" algorithm, a variation of SRC robust also to misaligned faces.

In this study we present an alternative solution, which takes advantage of the inner robustness of SRC in case of shift (and rotation); while doing so, we monitor also all critical parameters introduced in Chapter 3. That is, first of all we repeated the shift experiment, or, better a variation of it, because this time the test set stores also the original perfect aligned sample.

For what is concerning memory requirement, the size of the dictionary is 504 × (20 × 38) and the number of test samples is (3 × 38) in the 1st trail and (3 × 5 × 38) in all subsequent experiments.

Results are stored in Table 4.7, where columns “SL”, and “CL” show the median value of the sparsity and compressibility level, “DictCoh” is the coherence of the dictionary, “DictTestCoh” is the mutual coherence between the dictionary and the test matrix and “DeltaScore” is the confidence level of the classifier; all these parameters were introduced in Chapter 3.

Table 4.7 Blind experiment on shift

Number of shifted pixels	SRC (%)	SL (median)	CL (median *100)	DictCoh (*100)	DictTest Coh (median *100)	Delta Score (median *100)
0 pixels	100	65	96	99.76	99	96
2 pixels	100	64	94	99.76	99	86
4 pixels	100	61	92	99.76	99	68
5 pixels	100	60	90	99.76	98	58
6 pixels	98.42	59	89	99.76	98	47
8 pixels	81.05	57	87	99.76	97	16
10 pixels	60.35	56	86	99.76	96	14

In this case, both the median values of the sparsity level (SL) and the one of the compressibility level (CL) decrease together with the recognition rate of SRC; however, as pointed out in Chapter 3, we believe that only CL can be considered a critical parameter of SRC. Furthermore, the incoherence level of the dictionary is always the same, simply because the dictionary does not change, while the median values of *DictTestCoh* and *DeltaScore* start respectively from “0.99” and “0.96”, with SRC performance of 100%, to drop to “0.96”and “0.14” with SRC performance of 60%. As an example, let us consider the last row of Table 4.7: the dictionary is made up of perfectly aligned training samples while test samples are the original ones plus the ones shifted in all directions by 10 pixels; the recognition rate is not acceptable any more, and the medians of “CL” and “DictTestCoh” and “DeltaSCore” are the lowest of their columns.

At the same time, results of Table 4.7 suggest that a sparse classifier is robust up to 5 pixels shift, corresponding to a 3% shift of the original image; that is a dictionary of perfectly aligned faces can correctly classify test faces shifted up to 5 pixels. This observation gave the idea to use a shifted dictionary, as detailed in the following section.

4.3.5.1 **Shifted (and/or Rotated) Dictionaries**

Table 4.7 showed that sparse representation based classifier has the advantage to be robust to a limited amount of mis-registration due to shift (and rotation). The idea explored in this paragraph is to increase the robustness of SRC by using a shifted (and/or rotated) dictionary. The amount of shift tolerated by SRC depends to the robustness of the aligned dictionary, and, therefore, to the low dimension we are working with; in case of dimension 504, Table 4.7 showed that the aligned dictionary is robust up to 5 pixels shift and this suggests the use of a 5 pixels shifted dictionary, where the dictionary is made up of original and images shifted by 5 pixels, while test samples are shifted in all directions by different amounts of pixels. That is, in this experiment we use a 5 pixels shifted dictionary, meaning that out of every perfect aligned train sample the dictionary has now 5 atoms, the original one side by side to the 4 shifted versions (up, down, left, right).

For what is concerning memory requirement, the size of the dictionary is 504 × (20 × 5 × 38) and the number of test samples is (3 × 38) in the 1^{st} row and (3 × 5 × 38) in all other trials.

Table 4.8 Informed experiment on shift

Number of shifted pixels	SRC (%)	SL (median)	CL (median*100)	DictTestCoh (median*100)	DeltaScore (median*100)
0 pixels	100	62	98	99	94
2 pixels	100	62	98	99	94
4 pixels	100	62	98	99	94
5 pixels	100	63	99	99	96
6 pixels	100	61	98	99	93
8 pixels	100	58	98	99	79
10 pixels	100	57	97	98	59
12 pixels	88.82	54	96	98	30

Results of Table 4.8 shows that a shifted dictionary is robust up to 10 pixels shift in the original image, which is equivalent to 6% shift of the original image, a reasonable amount of misalignment produced by an average face detector. Moreover, Table 4.8 monitors the behaviour of the critical parameters of SRC by showing how “CL”, “DictTestCoh” and “DeltaScore” are correlated with the success rate of the classifier; the *DeltaScore* measurement is the most sensitive one.

Finally, because Table 4.6 showed that SRC is robust also to some rotation, it is possible to create an expanded dictionary of aligned and rotated faces capable of handling both shifts and rotations. Important to notice that, in case of a non-aligned database, the resulting dictionary is naturally shifted and rotated and this has the advantage to increase its robustness; such a situation is the one presented in Chapter 6 with the Extended Cohn-Kanade dataset.

CHAPTER 5

GEOMETRIC NORMALIZATION AND BLOCK-BASED APPROACH

The aim of this chapter is to describe the geometric normalization and the block based approach, which have been used together with SRC. That is, we wanted to determine the extent to which normalization of the input images could easier the classification task as well as we investigated the potentialities of SRC when used in a block based fashion.

5.1 Geometric Normalization

Faces in the Extended Cohn-Kanade (CK+) database [37], [38] are not aligned and, therefore, they may require geometric normalization. In addition they have background, which demands some form of cropping. By using different types of alignment and cropping we generated different experimental setups; in the following sections we give an overview of these experiments.

Cropped Faces

In the first approach we did not make any geometric normalization and we cropped all images using the face landmarks coordinates given in the database. The CK+ database comes with 68 landmarks associated with every face, which can be used for a simple cropping by selecting the most (left, right, upper, lower) landmarks, as shown in Figure 5.1:

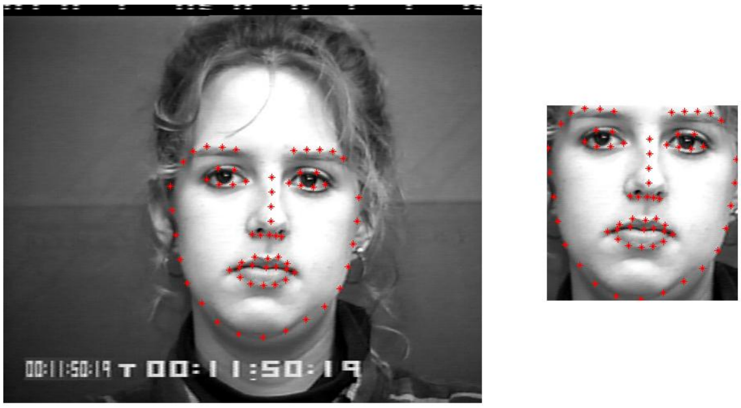


Figure 5.1 An original and a cropped face in the CK+ database

This non-specific cropping is admissible since experiments in Chapter 4 showed the robustness of SRC to mild shift and rotation, as well as the increased robustness of SRC with a shifted and rotated dictionaries. We had referred to this set of faces simply as ***Cropped Faces (CF).*** *A*s we can see from pictures in Figure 5.2, cropped faces have mild rotation, zoom, pose perturbations as well as some illumination effects.

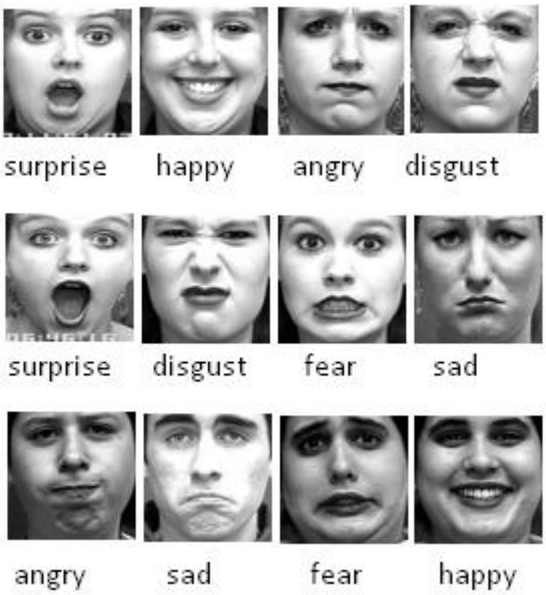


Figure 5.2 Cropped Faces

Cropped Faces with Geometric Normalization

We geometrically normalized all faces with respect to the location of the eyes; more in details, we considered the following points: (1) de-rotating the face to set the angle of the line passing through the pupils to "0" degrees, (2) scaling the face to impose a fix distance between the two pupils, and (3) imposing photometric normalization to deal with illumination changes. The following picture shows the sequence of steps:

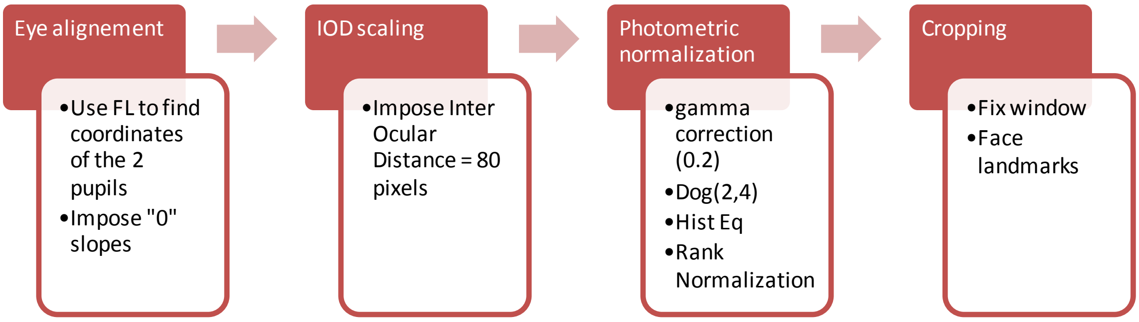


Figure 5.3 Geometric normalization steps

In order to rotate the face we used the data of the six face landmarks in the eye's region; that is, we averaged their coordinates to locate the left and right pupils of the eyes. The slope of the line passing through the eyes is the de-rotation angle, which allows for the eye alignment step, as shown in Figure 5.4:

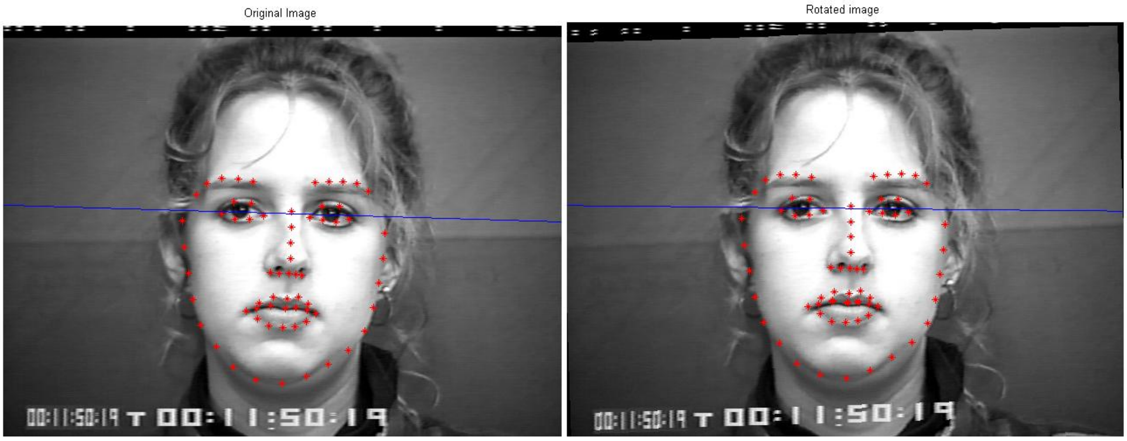


Figure 5.4 Original and rotated face

The second step of the geometric normalization process, is realized by imposing Inter Ocular Distance (IOD) equal to 80 pixels, which results is a shrink of the rotated image, as showed in the following picture:

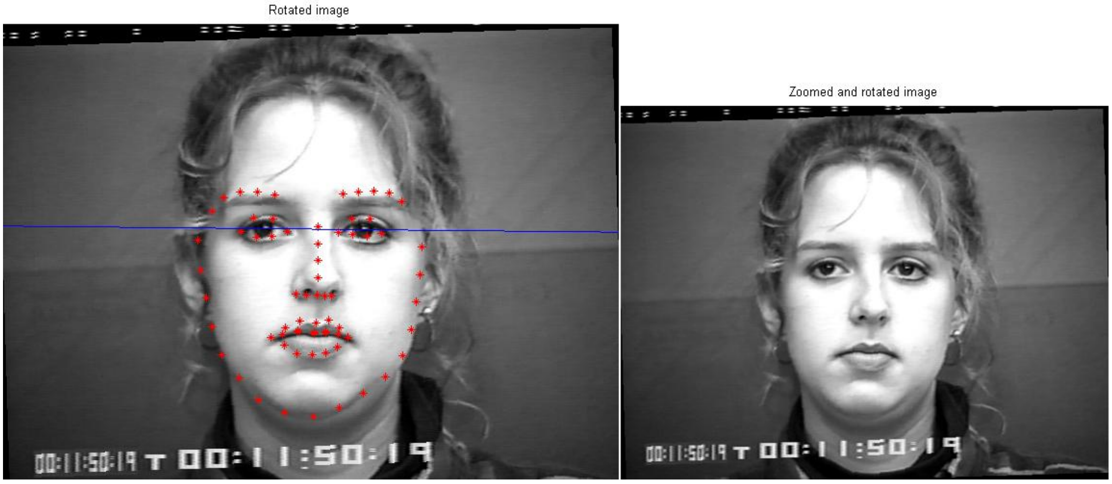


Figure 5.5 Rotated and zoomed face

Among all possible techniques available to achieve photometric normalization [39], we tested the Difference of Gaussian (DoG) filter with sigma values (2, 4); in some cases, we previously bighted the input picture with a gamma correction function so as to light the final results. Alternatively to the DoG filter, either we applied the histogram equalization function to all images or we imposed the same mean grey-level to all faces in the database, via rank normalization. The following diagram schematizes all tested paths:

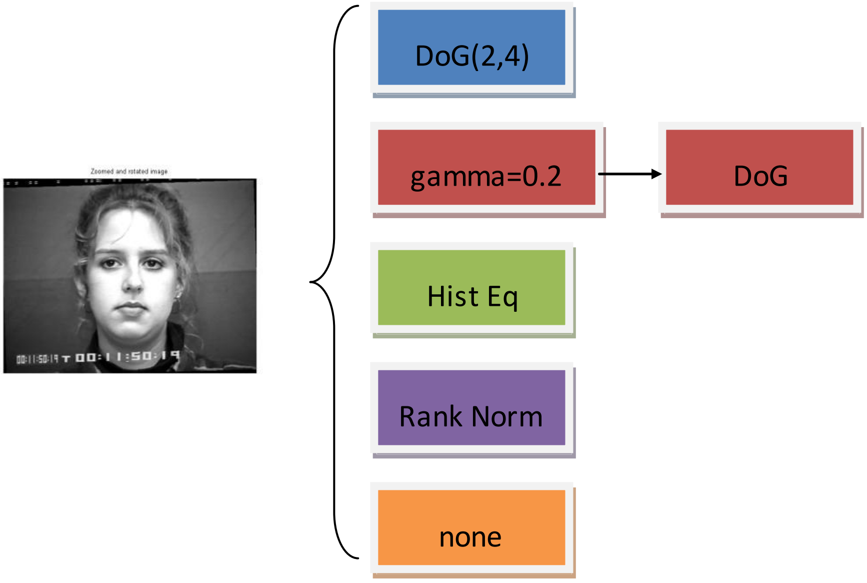


Figure 5.6 Photometric normalization

The aligned, zoomed, and filtered images need finally to be cropped, so as to separate every face from its background. After eye alignment and IOD scaling, it is common practice to crop the face with a rectangular window of fix size; we imposed (width, height)=(2.5*IOD, 3*IOD), which results in a 200×240 pixels window. In the following picture there are some of the faces cropped with this technique:

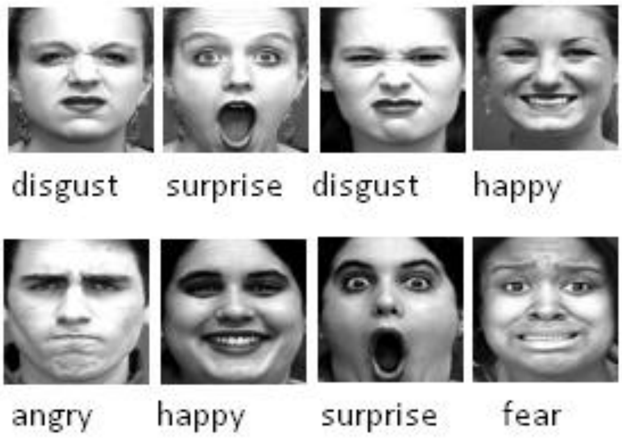


Figure 5.7 Aligned faces cropped with a fix size window

As we can see from Figure 5.7, in some cases the window is too tall while in other cases it is to short; in other words, the “fixed size window” is probably not the best technique to crop emotional faces, simply because different emotions change the size of the face. Alternatively, we used the rotated and zoomed face landmarks to crop the rotated and zoomed faces simply by selecting the most (left, right, upper, lower) coordinates to delineate the border of the cropping rectangular. Figure 5.8 shows some images of the new experimental setup.

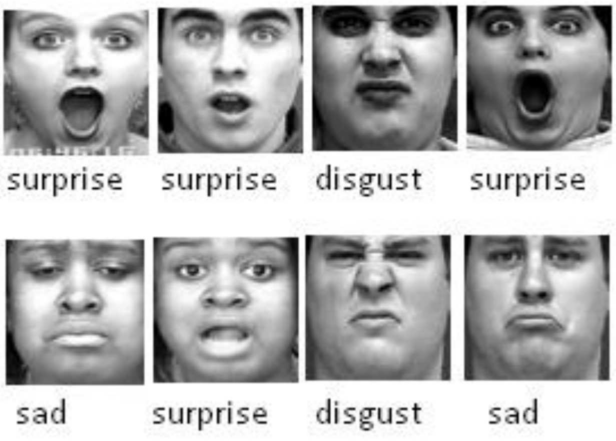


Figure 5.8 Aligned faces cropped with face landmarks

Important to notice that the eye-centered normalization and cropping technique has the advantage to produce perfectly aligned images, with respect to left and right pupils, but it may be too rigid to accommodate the different size variations on emotional faces. On the contrary, the bounding box cropping technique adapts itself to size variations of the face, but does not produce aligned faces, in other words, the coordinates of the left and right pupils will not be always the same. In both cases, the inner robustness of SRC can mitigate these mismatches.

In this study we refer to the set of faces with eye-centered alignment a ***Cropped Faces with Geometric Normalization (CFGN)***.

While the first two steps of “eye alignment” and “IOD scaling” are well defined, there are multiple possibilities to implement “photometric normalization” and “cropping”. This variety generates a number of experimental setups, some of them listed in the following table, together with their recognition rates:

Table 5.1 Performance of different experimental setups

	Eye Alignment	IOD	Gamma	DoG, DCT, HistEq, RankN	Window (W) Face Landmarks (FL)	Rec. rate
Cropped Faces	No	No	No	None	FL	86
CFGN1	Yes	Yes	No	DoG	W	86
CFGN2	Yes	Yes	No	None	W	83
CFGN3	Yes	Yes	No	DoG	FL (10 pixels)	70
CFGN4	Yes	Yes	No	None	FL	88
CFGN5	Yes	Yes	No	None	FL (10 pixels)	86
CFGN6	Yes	Yes	No	HistEq	FL (10 pixels)	76
CFGN7	Yes	Yes	Yes	DoG	FL (10 pixels)	55
CFGN8	Yes	Yes	No	DCT	FL	66
CFGN9	Yes	Yes	No	RankN	FL	88

Every row of Table 5.1 labels an experimental setup, the symbols “CF” and “CFGN” stand for “Cropped Faces” and “Cropped Faces with Geometric Normalization”. The header of Table 5.1 fixes the domains of the corresponding columns:

- “Eye alignment”, “IOD”, and “Gamma correction” may be either present or not;
- Possible values of column “DoG or DCT or HistEq or RankN” are {DoG, DCT, HistEq, RankN, none} so as to be able to specify the type of filter (if any) or pre-processing used in the current experimental setups;
- Column “Window (W) or Face Landmarks (FL)” may be read in a similar way even if, in this case, we cannot have the “none” value because some form of cropping is absolutely necessary, and we have the option to crop the face 10 pixels high with respect to the most upper face landmark coordinate as suggested by previous studies on action unit identification, where it was showed that successful experiments require a bit of front-head present in the cropped faces.
- Finally, the last column stores the recognition rate of the corresponding experimental setup.

The most successful case is for row CFGN4: cropped faces with eye alignment, IOD scaling, no filtering and cropping based on face landmark reaches the top performance of 88%, exactly the same as Lucey et al. with the much more complex AAM technique; in CFGN9 the introduction of rank normalization does not increase the performance.

Interesting to point out that the first row of the table, labelled CF, stores the recognition rate of the first experiment run with misaligned faces, it reaches the same recognition rate as CFGN1, which uses the DoG filter followed by a fix size window’s cropping. This is probably due to the inner robustness of SRC augmented by the use of a naturally shifted, rotated dictionary made up of faces with illumination changes, which makes it invariable to a little amount of in-plane rotation and illumination changes.

5.2 Block Based Classification

In a block based approach the original picture is divided into blocks, a first assignment of scores is performed block-wise and, finally, during the performance evaluation stage, all block's scores are fused together to allow for classification. Obviously the choice of the fusing technique is critical for the final performance. Picture 5.9 shows a cropped face divided into 2×2 and 3×3 blocks.

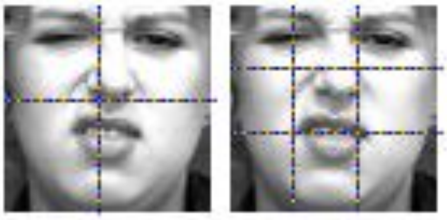


Figure 5.9 A face image partitioned into 2×2 and 3×3 blocks

In this experimental setup, every dictionary is three dimensional (3D), where the 3-th dimension is the block number, and every block is classified using the corresponding two-dimensional (2D) block-dictionary, a dictionary made up of blocks in the same position. Figure 5.10 shows the snapshot of the 3D dictionary used for emotion classification with the LOSO technique, when the test subject is subject1 who has only one test sample belonging to class 3:

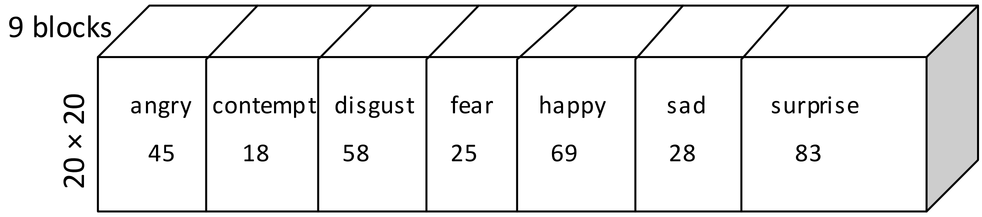


Figure 5.10 Snapshot of a 3D dictionary

Snapshot of Figure 5.10 refers to the case where we are working with 60×60 images divided into 3×3 blocks of size 20×20; the test subject has only one emotional face which results in a dictionary of 326 atoms, training faces; in a block based approach every image is divided into blocks and the first assignment of scores is block wise. Also the resulting matrix of scores is 3D, that is, every peak face has a set of scores associated to every possible class, as many scores as the number of used blocks; in order to be able to take the final decision it is, therefore, necessary to convert the score matrix from 3D to 2D; that is, classification requires to have only one score assigned to every couple (peak face, class). The following picture shows the conversion from a 3D to a 2D matrix of scores in case of a 3×3 block based approach:

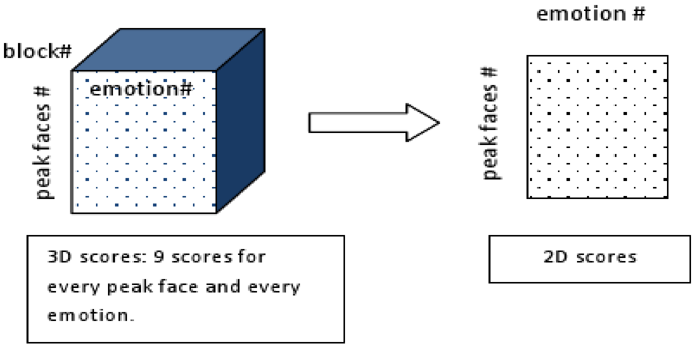


Figure 5.11 Conversion from a 3D to a 2D matrix of scores

The arrow in Figure 5.11 is implemented by a fusing technique, and, obviously, this conversion can be done in many different ways depending also to the problem under investigation. For example, in this study, we applied the block based approach to both emotion and action units (AUs) classification issues; while an emotion, which is made up of a collection of action units, is spread all over the face, on the contrary, a single AU is stored into only one or two blocks of the face, that is, it is a local features of the face. For example, in case of 3×3 AUs block based classification, it makes sense to divide AUs into “upper” and “lower” action units and fusing the scores by considering only blocks belonging to the top/bottom two rows. The obtained performance of this block based approach can probably be increased with the use of local masks as in [40]. On the contrary, in case of emotion classification, the useful information is spread all over the face, but, depending to the emotion under investigation, the true (unknown) class of the test sample, some blocks may have more discriminative power than other. The most successful fusing technique is the one that automatically identifies the most discriminative blocks and averages only those scores. That is, the block based approach of emotion classification is more challenging than the AUs one, because, for every emotion, it requires to automatically identify the collection of blocks storing discriminative information. This study focuses on the block based emotion classification experiment with a 3×3 block partition. The following paragraphs give an overview of the more interesting fusing techniques which were tested for this particular issue.

5.2.1 Fusing Algorithms

Average, Max, Min of all Block's Scores

Starting from a block based classification algorithm, we converted the score matrix from 3D to 2D simply by calculating the average among all scores or by selecting the maximum/minimum score for MCC/DFS.

Weighted Average of all Block's Scores

Starting from a block based classification algorithm, we converted the score matrix from 3D to 2D by assigning different weights to blocks in different positions. The following pictures shows the two masks used:

$$\frac{1}{16}\begin{bmatrix} 1 & 2 & 1 \\ 2 & 4 & 2 \\ 1 & 2 & 1 \end{bmatrix} \qquad \frac{1}{23}\begin{bmatrix} 3 & 3 & 3 \\ 2 & 3 & 2 \\ 2 & 3 & 2 \end{bmatrix}$$

Figure 5.12 Masks used for the fusing step

The logic of the first mask is to give more importance to the central block of the face, thinking that most of the emotion is concentrated around nose and mouth area; while the idea of the second mask is to give more importance to the blocks of the top row and those of the central columns of the face.

Other possible and interesting weights are related to (1) the success rate, (2) the amount of confidence of every block and (3) the compressibility level of the returned coefficient's vector. That is, in the first case, the weight assigned to every block is the success rate of that block in detecting that emotion, in formula,

$$P\_weight_b^{em} = \frac{P_b^{em}}{\sum_{i=1}^{block\_no} P_i^{em}} \tag{5.1}$$

$\forall b = 1, \cdots, block\_no, \forall em = 1, \cdots, emotion\_no,$ where P_b^{em} is the probability of success of block b in detecting emotion em. Obviously, this technique requires to partition the total number of samples into {train, validation, test} sets. In the second case, weights are the amount of confidence present in the decision step of every block, which is calculates as the distance between the score of the winner emotion with the one of the runner up. As said before, in case of emotion classification, the number and the position of discriminative blocks changes from emotion to emotion, and we are looking for an algorithm to detect those blocks automatically. We conjecture that classification is more successful when there is no uncertainty about the winner class and this happens whenever there is a big distance between the score of the winner and the one of the runner up; in order to cast the resulting weight into the interval [0,1], the obtained distance is normalized in the following way:

$$C\_weight_b^{test} = \frac{(score_{winner} - score_{runner\_up})}{\sum_{i=1}^{block\_no} score_i^{test}} \tag{5.2}$$

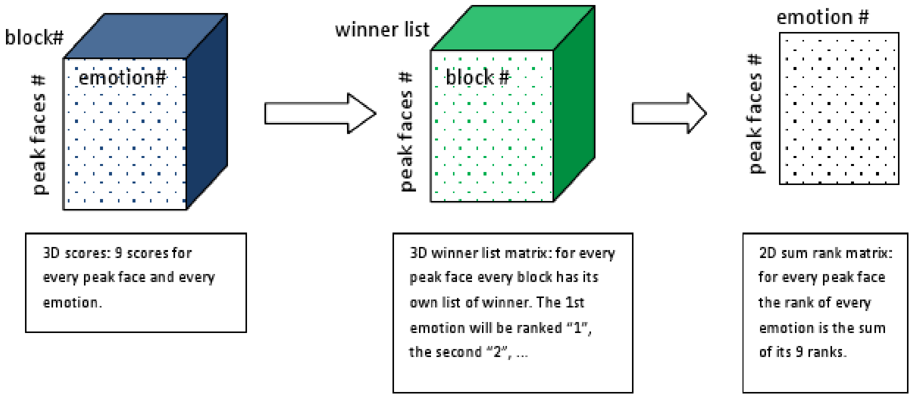
$\forall b = 1, \cdots, block\_no, \forall test = 1, \cdots, test\_no.$

Finally, we waited the score returned by every block with the compressibility level (CL) of the corresponding solution; this fusing technique was suggested by the results of Chapter 4, where the compressibility level of the coefficient vector looks to be correlated with the success rate of the current experiment.

Rank-based Fusing

Starting from a block based classification algorithm, we converted the block scores into ranks, from 1 to 7, and summed the rank scores over the blocks to get the final classification score. In order to do that, we converted the 3D score matrix into a 3D winner-list matrix, by exchanged the 2nd and 3rd dimensions while ordering the elements of the 3rd dim. The 3D winner-list matrix allows for ranking the emotions: for every peak face (327), every block (9) has a list of winner emotions sorted from the most to the least probable one. The algorithm creates the 2D sum-rank matrix by scanning the nine winner lists of every test face ranking them. The emotion that collects the lowest rank is declared as the emotion of the face.

Figure 5.13 gives the graphical representation of the algorithm:



3D scores: 9 scores for every peak face and every emotion.

3D winner list matrix: for every peak face every block has its own list of winner. The 1st emotion will be ranked "1", the second "2", ...

2D sum rank matrix: for every peak face the rank of every emotion is the sum of its 9 ranks.

Figure 5.13 Rank based algorithm

CHAPTER 6

SRC WITH EMOTIONAL FACES

The aim of this thesis is to explore the capabilities of the sparse representation based classifier. In this chapter we focus on the problem of 2D face classification with emotional faces; faces with expressions present two inter-related problems: the first one is the automatic classification of the emotion, as reflected on the facial expression, and the second challenge is the identity recognition task with emotions; that is, we are interested to investigate the amount of deterioration caused by different expressions. We addressed both issues based on the paradigm of sparse coding; the study is conducted experimentally using the Extended Cohn-Kanade dataset [37], [38]. More specifically:

- We explored automatic recognition of emotions and the detection of their constituent elements, e.g. action units (AUs), within the framework of the Facial Action Coding System (FACS) [41];
- We analysed the biometric performance of a sparsity based algorithm vis-á-vis emotion categories. The goal is to determine the degree to which different categories of facial expressions affect correct identification.

For every experiment we reported its success rate together with its memory requirement, as a measure of its complexity.

6.1 Classification of Emotional Faces via SRC

The presence of facial expressions may not be the top factor handicapping reliable facial identity recognition as compared to pose, illumination and occlusion. However, it still deserves some attention as it affects negatively especially in case of strong expressions. Expression detectors can be useful in the design of intelligent human-computer interaction systems, in developing effective interfaces, or to enhance man-machine communication. Next generation human-computer interfaces on robots are expected to understand non-verbal communication clues of humans and to decode the perceived mental state images into contextual information, [42].

We used a sparse classifier to tackle the issues of emotion classification, action units (AUs) identification and person identification in the presence of expressions. In case of emotion classification, we competed with the benchmark protocol given by Lucey at al. in [38]; we then used the same technique to provide a solution for AUs identification, and, finally, we considered the problem of subject identification despite emotion: we investigated the robustness of the sparse classifier in identifying a subject with emotion, using only other emotional faces; the problem is first faced with a mono-class dictionary, which is then expanded to a multi-class one.

6.2 Description of the Database

We worked with the Extended Cohn-Kanade (CK+) dataset [37], [38], which is one of the few databases with coded emotions; it is a video clip dataset with validated emotion labels. The database has 123 subjects, and for every person, there is a variable number of sequences, from 1 to 13. Every sequence is made up of a different number of frames, from 3 to 71, all starting from a neutral face and reaching the peak expression. There are 593 video clips (#Subjects ×#Sequences), and thus not all subjects act all emotions, and emotions (which are classes) have, therefore, differing number of samples. More in detail, for every sequence there is Facial Action Coding System (FACS) metadata file, containing the list of AUs present in the apex frame. Lucey et al., [38], designed a three step selection procedure using the Emotion Prediction Table of [41] to associate a validated emotion code to apex face images.

This resulted in 118 subjects having validated emotion labels, and only 327 out of 593 sequences having a validated coded emotion.

6.3 Experiments

Faces in the database are not aligned nor cropped, and they have differences in illumination. In the first approach, we cropped the faces based on the landmark coordinate data provided with the database, that is, for every picture, we selected the extreme points, (left-most, right-most, upper-most, lower-most) coordinates to delineate the borders of the cropping rectangle. The resulting cropped faces have different sizes and require the use of some dimensionality reduction methods. Considering the results of Paragraph 4.3.1, we decimated all images to 64×64 pixels. Notice that the dictionary built as concatenation of vectorized faces is not necessarily over-complete, and that the resulting cropped faces are not aligned; however, no effort was made to compensate for their shift, rotation, zoom effects, nor for their substantial illumination changes. In the second approach the images are first geometrically normalized before being cropped and resized to dimension 64×64. In the sequel we will refer to these two approaches simply as ***Cropped Faces (CF)*** and ***Cropped Faces with Geometric Normalization (CFGN)***, meaning that, if not specified, the geometric normalization is not present. While the second approach is the common one, the idea of working directly with misaligned faces comes from the previous experiments on geometric distortions (Paragraph 4.3.5), where we tested the robustness of SRC to a limited amount of shift and rotation. Moreover, in the same section we showed that a dictionary enriched with controlled shift and rotation becomes, as expected, more robust that the original one and this is the natural product of the first approach, the Cropped Faces; in other words a database of misaligned faces produces a misaligned dictionary which is robust to shift and rotation.

6.3.1 Emotion Classification

For emotion classification, we repeated the same experiment given as benchmark in [34] with Leave-One-Subject-Out (LOSO) cross validation technique. Having an initial dataset of 45 Angry (An), 18 Contempt (Co), 59 Disgust (Di), 25 Fear (Fe), 69 Happy

(Ha), 28 Sadness (Sa) and 83 Surprise (Su) faces belonging to 118 different subjects, the leave-one-subject-out technique allows for 118 trials using 327 video clips. We worked only with apex frames having a validated emotion label. For every subject, we created a dictionary, D=[An|Co|Di|Fe|Ha|Sa|Su], made up of peak faces of the remaining 117 subjects.

For what is concerning memory requirements, the 1st dimension of the dictionary is fixed to 64 × 64 while the 2nd dimension changes slightly from individual to individual, depending on the number of emotional faces of the current subject. For example, the 1st subject has only 1 emotional face hence the dictionary is 4096×326, while subject 2 has 3 emotional faces and the dictionary is 4096×324.

We formulated emotion recognition as 7-class problem, that is:

$$\arg\min_i \| y - D \cdot (0, \cdots, x_1^i, \cdots, x_{m_i}^i, \cdots, 0) \|_2 \tag{6.1}$$

where m_i is the cardinality (number of columns) of the i-th emotion class in the dictionary. The following picture shows an example of dictionary used for emotion classification:

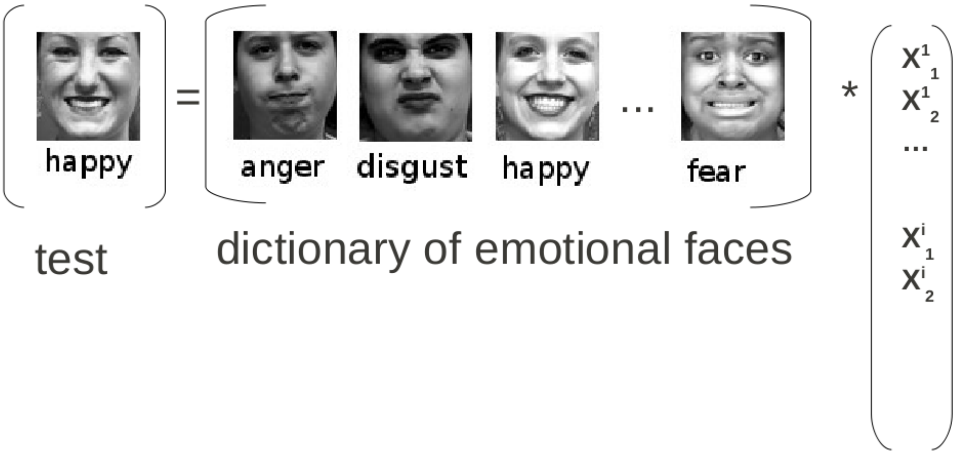


Figure 6.1 The emotion classification problem

As a comparison to our study, the method in [38] is based on the Active Appearance Model (AAM) paradigm. In contrast to the simplicity of the SRC method, the AAM-based emotion classification is considerably more complicated. Briefly, it consists of the following steps:

1. After an initial set of manually marked faces, other faces are landmarked using a gradient descent fitting of the AAM
2. A base shape is obtained using Procrustes analysis
3. A set of rigid shape parameters vis-á-vis base shape are extracted using the Principal Component Analysis of the coordinates of 68 landmarks on the face
4. A second set of shape parameters are obtained after all non-rigid shape variations are compensated for using piecewise warps on triangle patch appearances
5. Separate one-against-all SVMs are trained for each emotion, and their scores are fused using logistical linear regression.

In the sequel, we will refer to the two methods simply as SRC and AAM.

Working with the Whole Image

Since the CK+ database contains seven emotions, there are seven compartments in the dictionary; first we worked with the whole faces, that is, we created a dictionary where every column is a vectorized form of the entire face. At every run, all apex faces of the coded sequences of one subject are used for testing against a dictionary made up of apex faces of annotated sequences of the remaining 117 subjects. The performance of the SRC algorithm with misaligned cropped faces is 86%, while the use of geometric normalization increases the recognition rate up to 88%, exactly like in [38]. Table 6.1 and Table 6.2 show the confusion matrices of SRC when working with cropped faces (CF) and with cropped faces with geometric normalization (CFGN): rows label the true classes, columns correspond to the predicted classes, while the cell values are the recognition rates, in percentage; SRC results are compared with the best performance of Lucey et al. (in brackets) obtained with AAM technique.

Table 6.1 Confusion matrix of holistic SRC with CF (benchmark results)

Recognition Rate	An	Co	Di	Fe	Ha	Sa	Su
An	**71.0 (75.0)**	4.0 (5.0)	18.0 (7.5)	0.0 (5.0)	0.0 (0.0)	4.0 (5.0)	2.0 (2.5)
Co	11.0 (3.1)	**72.0 (84.4)**	6.0 (3.1)	6.0 (0.0)	6.0 (6.3)	0.0 (3.1)	0.0 (0.0)
Di	5.0 (5.3)	0.0 (0.0)	**92.0 (94.7)**	0.0 (0.0)	0.0 (0.0)	2.0 (0.0)	2.0 (0.0)
Fe	4.0 (4.4)	12.0 (8.7)	0.0 (0.0)	**64.0 (65.2)**	12.0 (8.7)	4.0 (0.0)	4.0 (13.0)
Ha	0.0 (0.0)	0.0 (0.0)	0.0 (0.0)	0.0 (0.0)	**100.0 (100.0)**	0.0 (0.0)	0.0 (0.0)
Sa	14.0 (12.0)	4.0 (8.0)	7.0 (4.0)	4.0 (4.0)	0.0 (0.0)	**68.0 (68.0)**	4.0 (4.0)
Su	2.0 (0.0)	1.0 (0.0)	0.0 (0.0)	0.0 (0.0)	0.0 (0.0)	2.0 (4.0)	**94.0 (96.0)**

It is interesting to notice that the happy emotion is the easiest to be recognized, and this means that a happy face is never confused with another expression. The second most recognizable emotion is surprise with a correct score of 94.0%; this is expected because the open mouth present in surprised faces probably maps that class far away from the others. Finally we observe that the problematic expressions are: (1) contempt and sad emotions often confused with the anger, (2) fear expression misinterpreted as contempt or happy, and (3) angry faces confused with disgusted ones.

Table 6.2 Confusion matrix of holistic SRC with CFGN (benchmark results)

Recognition Rate	An	Co	Di	Fe	Ha	Sa	Su
An	**80.0 (75.0)**	2.0 (5.0)	11.0 (7.5)	2.0 (5.0)	5.0 (0.0)	0.0 (5.0)	0.0 (2.5)
Co	11.0 (3.1)	**78.0 (84.4)**	6.0 (3.1)	0.0 (0.0)	5.0 (6.3)	0.0 (3.1)	0.0 (0.0)
Di	7.0 (5.3)	0.0 (0.0)	**90.0 (94.7)**	0.0 (0.0)	1.0 (0.0)	0.0 (0.0)	2.0 (0.0)
Fe	8.0 (4.4)	8.0 (8.7)	0.0 (0.0)	**76.0 (65.2)**	4.0 (8.7)	4.0 (0.0)	0.0 (13.0)
Ha	0.0 (0.0)	0.0 (0.0)	1.0 (0.0)	0.0 (0.0)	**99.0 (100.0)**	0.0 (0.0)	0.0 (0.0)
Sa	7.0 (12.0)	7.0 (8.0)	10.0 (4.0)	1.0 (4.0)	0.0 (0.0)	**71.0 (68.0)**	4.0 (4.0)
Su	0.0 (0.0)	4.0 (0.0)	0.0 (0.0)	0.0 (0.0)	0.0 (0.0)	1.0 (4.0)	**95.0 (96.0)**

Also with geometric normalized faces, happy and surprised emotions have the highest recognition rates, while angry, contempt and sad faces are reciprocally confused with disgusted, angry and disgusted faces.

In order to increase the current performance, we run different experiments (1) substituting vectorization with zig-zag coding, (2) checking for different size of the images (up to 124*124, after that there is “out of memory” problems), (3) different types of cropping (including Viola-Jones), (4) changing the number of max iterations in the synthesis algorithm, (5) considering the last two peak faces out of every coded sequence, and (6) trying some robust versions of SRC (introduced in Chapter 2), but we did not get any better. Finally, we investigated the potentialities of a block based approach, as suggested in [43] and [44].

A Block Based Approach

The aim of this experiment is to increase the success rate of the holistic SRC on misaligned cropped faces; the initial performance of this experiment is 86%, Table 6.1 shows the corresponding confusion matrix. In a block based approach we divided the cropped and decimated (not aligned) images in a 2×2 fashion, which results in blocks’ size of 30×30. Table 6.3 shows the confusion matrix of the block-based SRC: as usual, rows label the true classes, columns correspond to the predicted classes, while the cell values are the corresponding percentages; SRC results are compared with the best performance of Lucey et al. (in brackets) obtained with the AAM technique.

Table 6.3 Confusion matrix for 2×2 blocks based SRC on CF (benchmark results)

Recognition Rate	An	Co	Di	Fe	Ha	Sa	Su
An	**80.0** **(75.0)**	4.0 (5.0)	9.0 (7.5)	0.0 (5.0)	0.0 (0.0)	4.0 (5.0)	2.0 (2.5)
Co	6.0 (3.1)	**67.0** **(84.4)**	0.0 (3.1)	6.0 (0.0)	6.0 (6.3)	11.0 (3.1)	6.0 (0.0)
Di	2.0 (5.3)	0.0 (0.0)	**92.0** **(94.7)**	0.0 (0.0)	3.0 (0.0)	0.0 (0.0)	3.0 (0.0)
Fe	0.0 (4.4)	0.0 (8.7)	4.0 (0.0)	**64.0** **(65.2)**	20.0 (8.7)	0.0 (0.0)	12.0 (13.0)
Ha	0.0 (0.0)	0.0 (0.0)	0.0 (0.0)	0.0 (0.0)	**100.0** **(100.0)**	0.0 (0.0)	0.0 (0.0)
Sa	21.0 (12.0)	0.0 (8.0)	0.0 (4.0)	0.0 (4.0)	0.0 (0.0)	**71.0** **(68.0)**	7.0 (4.0)
Su	0.0 (0.0)	1.0 (0.0)	0.0 (0.0)	0.0 (0.0)	0.0 (0.0)	1.0 (4.0)	**98.0** **(96.0)**

Again happy and surprised faces have the highest recognition rate; in this case the main problem is with fearful and sad emotions too often confused with happy and angry expressions. The average performance of the block based SRC algorithm on misaligned faces is 88%, exactly the same as Lucey et al. We conjecture that there could be room for improvement by using overlapping blocks, better alignment (presently none), better cropping, setting the resolution optimally etc.

6.3.2 Action Units Identification

The aim of this experiment is to investigate the capability of SRC to identify action units (AUs); we worked both with the entire image and with image blocks. Like in [38] the experiment was run using the leave-one-subject-out (LOSO) cross validation technique, which allows for 123 trials; however, Lucey et al. worked with only manually FACS coded apex frames while we had to use all peak faces. Out of the total 64 AUs, we selected 16 AUs that occurred abundantly, in at least 50 apex faces. Thus the resulting dictionary has (16+1) classes, one per relevant AU plus an 'extra' class containing peak faces with AUs other than the 16 selected target ones, and the total number of AUs instances is 1988 distributed across the 16 relevant AUs, as in Table 6.4, notice that, since we considered all apex frames, the frequency of AUs is slighter higher than the one reported by Lucey et al.:

Table 6.4 Population of selected AUs

AU1	**AU2**	**AU4**	**AU5**	**AU6**	**AU7**
175	117	194	102	123	121
AU9	**AU12**	**AU15**	**AU17**	**AU20**	**AU23**
75	131	95	203	79	60
AU24	**AU25**	**AU26**	**AU27**		
58	324	50	81		

Working with the Entire Face

At every trial, all apex frames of the current subject are tested against a dictionary made up of peak faces of the remaining 122 subjects. Since every apex frame has more than one AU, it belongs to more than one class, and, therefore, the same training picture may be present with more than one occurrence in the dictionary. For the same reason the output of SRC will not have only one winner class, but an ordered list of candidate AUs; a success occurs whenever all present AUs are at the top of the list. Figure 6.2 gives the graphical representation of such a situation:

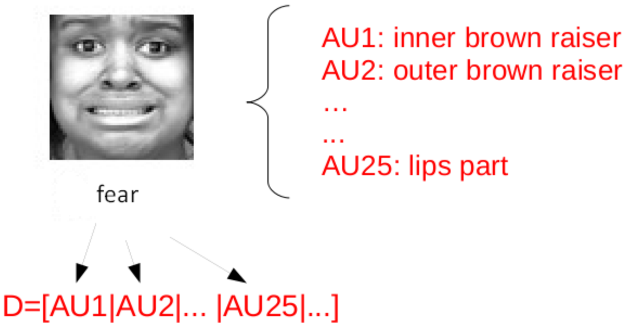


Figure 6.2 Emotional face with multiple AUs

The AU recognition algorithm is similar to the previous experiment where instead we solve a 17-class problem. To assess the performance of this experiment, we considered a modification of the Cumulative Match Count (CMC) technique, which considers a success whenever the correct class is the most probable one, and returns the total number of correctly classified classes. Considering the characteristics of this experiment, we decided to evaluate the performance of SRC by looking at the top *N* classes, where, for every peak face, *N* is the minimum between the number of AUs present in the current apex frame and the number of AUs that have received a non-negligible score; that is *N* is the dynamic threshold which varies from face to face. The modified CMC compares the list of present AUs, in the current test sample, with the list of the top *N* ones, and a success occurs whenever a true AU on the face is also in the detected set. Considering only the AUs selected in Table 6.4, the 593 peak faces of CK+ have a total of 1988 AUs, and the average recognition rate is therefore calculated by dividing the number of AUs that appear among the top *N* AUs (rank ordered according to their SRC scores) by 1988. To clarify this point, we remark that we know the correct type and number of AUs in the face, any present AU ranking lower than this number is considered as a mis. Figure 6.3 shows as the total number of correct classified AUs increases together with the flexibility. The N+5 tier recognition rate is 53%, which is already a good result to be improved with a block based approach.

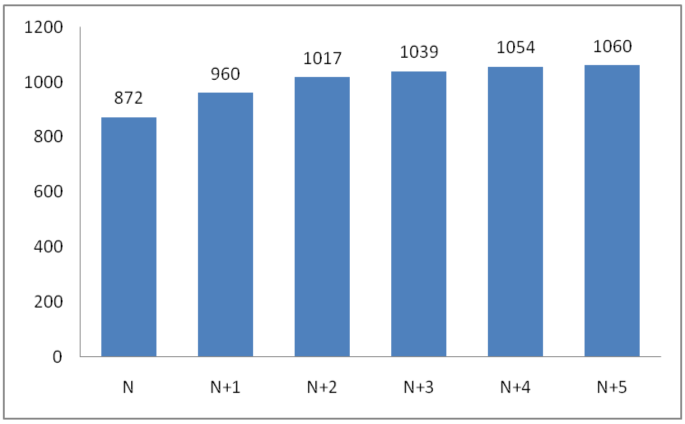


Figure 6.3 CMC scores for AUs identification on the entire face

A Block Based Approach

In a block based approach we divided the cropped and decimated images in a 3×3 fashion, which results in blocks of size 20×20. We ran the same experiment as before but over the blocks; that is, we applied the LOSO cross validation technique and we built a different dictionary for every subject. When working in a block based fashion every dictionary is 3D, where the 3-th dimension is the block number. As a result, every peak face has a set of 9 scores associated to every action unit. For performance evaluation, these 9 scores can be fused in various ways to calculate the probability value to be associated to every (peak face, AU) couple. The 1st fusing technique calculates the average of all scores, while the 2nd one distinguishes between "upper" and "lower" AUs and considers only blocks belonging to the top/bottom two rows in case of upper/lower AUs.

Figure 6.4 compares the resulting performance of the three different approaches in case of CMC(N+5): the whole face performance of 53% increases to 71% with a block-based approach calculating the average of all scores, and reaches 75% with a block-based algorithm based on upper/lower AUs separation.

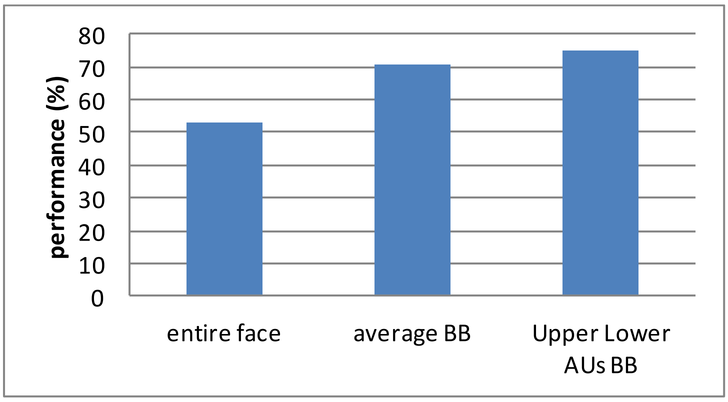


Figure 6.4 CMC scores at N+5 for AU identification. Bars left to right: whole face, average 3×3 blocks, upper/lower AUs 3×3 blocks

6.3.3 Identification Despite Emotions

The aim of this experiment is to compare the levels of classification difficulty that emotional faces cause. Thus pairs of emotions from among neutral, angry, contemptuous, disgusted, fearful, happy, sad, surprised faces were selected such that the dictionary was made up of one emotion, emotion1, and tested with another emotion, emotion2. Because not every subject has all emotions, it was first of all necessary to find out the total number of subjects having both emotion1 and emotion2. These pairs are shown in Table 6.5, where values denote the number of subjects present in both classes. Because the least represented emotion is contempt with only 18 cases, we set the threshold to 18 so that Table 6.5 shows in bold any couple with 18 or more representative pairs; obviously this table is symmetric with respect to the diagonal cells because the number of subjects having both (emotion1, emotion2) faces is the same as the number of subjects with (emotiona2, emotion1) faces.

Table 6.5 Distribution of the number of subjects sharing emotion pairs

	Ne	An	Co	Di	Fe	Ha	Sa	Su
Ne	**118**	**45**	**18**	**59**	**25**	**69**	**28**	**83**
An	**45**	**45**	4	17	11	**26**	13	**32**
Co	**18**		**18**	0	4	1	5	1
Di	**59**			**59**	13	**41**	12	**50**
Fe	**25**				**25**	13	8	**19**
Ha	**69**					**69**	**19**	**59**
Sa	**28**						**28**	**20**
Su	**83**							**83**

Out of every sequence with emotion, we selected the first 2 neutral faces and the last 2 peak-emotional pictures. Only emotion pairs for which at least 18 samples were available (contempt faces have 18 samples only) were considered; this explains the lacunae in Table 6.6 which stores the recognition rate, in percentage, of SRC; lacunae corresponds to cases where there was not an adequate number of samples.

Table 6.6 Subject recognition rate (percentage) across expression

SRC performance	Ne	An	Co	Di	Fe	Ha	Sa	Su
Ne	100	100	97	91	96	88	96	77
An	100	100				96		81
Co	97		100					
Di	90			100		93		64
Fe	96				100			79
Ha	93	94		94		100	87	71
Sa	95					97	100	78
Su	63	47		59	71	46	80	100
Mean	92	85	99	86	89	87	91	79

More in details, in Table 6.6, diagonal cells store the performance using faces of the same emotion in both train and test sets; obviously, in every trial the tested face has been removed from the training set; the recognition rate is always 100%. This is a very high score it could be deceptive because there are images from the same scene of the same person. The cells of the first row store the performance of a dictionary of neutral faces in recognizing faces exhibiting other emotions; notice that the lowest recognition rate occurs with the surprise emotion, and this is somewhat expected because the open mouth present in surprise faces probably maps that object far away from its class. The cells of the first column store the performance when dictionaries of emotional faces (a different emotion for every row) is used to classify neutral faces; again angry, contemptuous, fearful and sad faces have a success rate nearby 100%, while the worse performance is with surprised faces. Other cells can be interpreted in the same way.

For what is concerning memory requirement, the 1st dimension of the dictionary is fixed to 4096, because we are working at low resolution 64 × 64, while the 2nd dimension depends to the number of subjects sharing both emotions. For example, in case of (Ne, An), the 2nd experiment of the 1st row, the dictionary has 45 × 2 atoms for a total of 45 × 2 test samples, because we selected the first-last 2 neutral-peak faces out of every sequence coded with the angry emotion.

Finally, we attested the performance of richer dictionaries made up of an assortment of all expressions, except the tested one; that is the training set contains all other emotion classes except the one of the test sample. Table 6.7 stores the recognition rate of SRC.

Table 6.7 Subject recognition rate (percentage) with a richer dictionary

SRC performance (%)	Test
All-Ne	98
All-An	100
All-Co	100
All-Di	98
All-Fe	96
All-Ha	100
All-Sa	100
All-Su	83
Mean	97

Comparing the performance of Table 6.7 with the "mean" row of Table 6.6, we observe, not surprisingly, that a bigger dictionary, containing an assortment of emotional faces, is more successful, in general, than the one-emotion dictionary; e.g. the recognition rate of fearful faces passes from 89% to 96%, but the performance of surprised faces is still the worse one.

CHAPTER 7

CONCLUSIONS AND FUTURE WORK

7.1 Conclusions and Future Work

This thesis investigates the capabilities of the sparse representation based classifier, which proposes a new approach to the classification issue. The study focuses on face classification, because face remains the holy grain of biometric recognition and face recognition algorithms have achieved reasonably high levels of accuracy under controlled conditions; however, under non-ideal, uncontrolled conditions, as often occurs in real life, their performance is still very poor. The problem is very interesting, due to the big demand of accurate face recognition algorithms from fields such as security, for access control and surveillance task, and robotics, to enhance human-computer interfaces. That is, next generation recognition systems will need to identify people in real time and in much less constrained situations. We tested the performance of the SRC classifier under different disturbance elements, such as low enrolment size, presence of noise, illumination changes, geometric distortions, and expressions.

Another important contribution of this thesis is the study of some critical parameters of SRC, which are considered to be bound to its performance, and the proposal of new measurements, which revealed to be correlated to the success rate of SRC. The initial theoretic analysis of those critical parameters is then empirically supported by their monitoring along different experiments. That is, in this thesis we argue that the incoherence of the dictionary does not affect the success rate of SRC and, on the contrary, because a test sample can be represented as a linear combination of its class training samples only “if” and “when” its class is well represented, the classification rate can generally be increased by enlarging the dictionary and passing from a blind to an informed experiment. Moreover, as a critical parameter of SRC, instead of the level of sparsity we propose to consider the level of compressibility of the coefficient vector; and, finally, we introduced the idea of the level of confidence of the classifier, which turned out to be the most sensitive critical measurement of SRC.

To summarize, we have investigated the robustness of SRC, which is a new non linear face classifier based on sparse approximation, in different experimental setups. Our main conclusions can be summarized as follows:

1**. Experimental results on the Extended Yale B** database show that the SRC algorithm is superior to Fisher Linear Discriminant under all the adverse conditions tested. This implies that a classifier method based on sparse representation, in fact a generalization of the nearest neighbour method, is better than the well-known parametric method, like Fisher Discriminant Analysis. Our experiments show that:

- Resolution: The performance of SRC for decimated images with factor 24 is still 2 points better than the one of Fisher;
- Gallery size: For gallery size above 30, SRC and Fisher reach almost perfect recognition for the Extended Yale B database. For gallery sizes at and below 15 SRC outperforms Fisher classifier by at least 10 points;
- Noise: also in this case SRC behaves always better than Fisher. More in details, the initial 3 percentage points of difference in performance grows up to 75 percentage points with noisy pictures;
- Illumination: SRC outperforms Fisher algorithm in all experimental setups;
- Geometric distortions: SRC outperforms Fisher also in this case, even if such a low performance is not acceptable for both classifiers. In case of SRC the use of the shifted (and/or rotated) dictionary solves the problem.

One advantages of the SRC method is that it does not need any training, like the NN method, which makes it simple to implement. The price to pay for this simplicity is an increase in the testing time; in fact for the Ext. Yale B database with the gallery size of 35, SRC runs in 292 seconds while Fisher needs 74 seconds. Finally, sparse classifiers have the advantage to work well also with basic subspaces, such as down-sampled or random projected images.

2. **Experimental results on the Extended Cohn Kanade database** show that the robustness inherent in the sparse approximation algorithm has useful consequences in identification of emotional faces as well as AU and emotions themselves. The main conclusions to be drawn are:

- Emotion classification: When SRC tries to reconstruct a face with a given emotion from a pool of faces having various emotions, 88% of the time it selects faces from the same emotion class. This result stands favourably vis-á−vis the benchmark protocol given in [38] where the simplicity of SRC is in sharp contrast with the complexity of the landmark and AAM-based methods.
- AUs identification: Adapting the SRC method used for emotion classification to AU identification, working with the whole face, we obtained a promising performance figure of 53%, which is then increased to 71% with a simple block-based SRC and reaches 75% with a weighted block-based SRC.
- Subject identification in the presence of emotions: We investigated the degree to which mismatched emotions between training and testing affect correct identification. We ran both one emotion-to-another emotion tests as well as one emotion-to-the rest of emotions tests. While within-class identification (no mismatch) were nearly ideal, recognition rates suffer variably, and not surprisingly, proportional to the deformation that the emotion induces in the face. For example, surprise causes the biggest drops in performance. Richer dictionaries where all the rest of emotions are present do much better.
- Geometric normalization: following the standard practice we aligned all faces of CK+ before classifying them and we increased the original performance of (only) “2” percentage points. This little improvement is probably due to the inner robustness of SRC which can handle (not so) little amount of in-plane rotation and illumination changes. Especially the use of filters and histogram equalization was

particularly harmful for the subsequent classification step, probably because they alter the distinguishing characteristics of every face; while SRC does not require frontal illuminated images, due to its inner robustness.

- Block based classification: we wondered if a block based SRC could increase the recognition rate of the holistic one also in case of faces without occlusions. Our experimental results get little improvements but they worked with the original block-dictionary; that is, we did not want to expand the dictionary like in [44] by adding all shifted blocks due to complexity reasons. Moreover, we did not use overlapping blocks, yet.

3. **"Last but not least"** we made some **statistics studies** to justify the excellent performance of SRC and to identify the parameters which affect its performance; that is, we investigate "if" and "how" the sparse coefficients returned by a synthesis algorithm store enough discriminative information to enable classification of signal without reconstruction. The same question is addressed by Daventport et al. in [45] and answered from a theoretic point of view; our approach is "dual" because we started with empirical studies, searching for a theoretical justification into related statistical parameters. We are now confident in saying that the coherence of the dictionary is a requirement for the projection or analysis step; that is, the dictionary used to make random projection must be incoherent so as to preserve the distance between any two different faces, points in $\Re^N$; for this use, Gaussian random matrices are a possible choice. On the contrary, for sparse classification, the dictionary D is used only during the synthesis step, it is highly coherent because made up of faces, which are similar vectors, and it has different performance, depending on the input test sample. Finally, critical parameters capable of predicting the recognition rate of SRC are (1) the level of compressibility of the coefficients' vector, (2) the mutual coherence between dictionary and test sample; that is, the nearest is the test image to the dictionary the highest is the performance of SRC, and (3) the confidence of the classifier, which can be measured as the difference between the score of the winner class and the one of the runner-up.

Future Work

There are several avenues of research as a follow-up. Obviously the performance of the classifier is affected by the type of database; hence first we are going to consider the reproducibility of these results over alternate databases, like CMU PIE, FRGC, MMI, and Bogazici. Among the possible issues to be addressed there are:

- Using overlapping blocks with Sequential Floating Forward Search (SFFS)
- Testing SRC for face landmark detection, age progression, gene selection
- Challenging the 3D face recognition issue
- Dictionary Learning.

REFERENCES

[1] Pentland, A., and Choudhury, T., (2000) “Face Recognition for Smart Environments”, IEEE Computer”, 33 (2): 50-55.

[2] Jain, A. and Kumar, A., (2010). Biometrics of Next Generation: an Overview, Second Generation Biometrics, Springer (2010)

[3] Gong, S., McKenna, S.J., and Psarrou, A. (2000). Dynamic Vision: from Images to Face Recognition, Imperial College Press.

[4] Adini, Y., Moses, Y., and Ullman, S., (1997). “Face Recognition: the Problem of Compensating for Changes in Illumination Direction”, IEEE Transactions on Pattern Analysis and Machine Learning, 19: 721-732.

[5] OToole, A.J., Jiang, F., Roark, D., and Abdi, H., (2006). “Predicting Human Performance for Face Recognition”, Face Processing: Advanced methods and models, W-Y. Zhao and R. Chellappa, Eds. Elsevier.

[6] Calder, A.J., and Young, A. W., (2005). “Understanding the Recognition of Facial Identity and Facial Expression”, Nature Review Neuroscience, 6(8): 641-651.

[7] Barlett, M. S., Littlewort, G.C., Frank, M.G., Lainscek, C., Fasel I.R. and Movellan, J. R., (2006). “Automatic Recognition of Facial Actions in Spontaneous Expressions”, Journal of Multimedia, 1(6): 22-35.

[8] Gong, D., Yang, Q., Tang, X., and Lu, J., (2005). “Extracting Micro-Structural Gabor Features for Face Recognition”, Int. Conference Image Processing (ICIP), 11-14 September 2005, Genoa, Italy

[9] Wagner, A., Wright, J, Ganesh, A., Zhou Z., and Ma, Y., (2009). “Towards a Practical Face Recognition System: Robust Registration and Illumination by Sparse Representation”, IEEE Conference on Computer Vision and Pattern Recognition, Miami, FL, 597-604.

[10] Yan, S., Wang, H., Liu J., Tang, X., and Huang, T.S., (2010). “Misalignment-Robust Face Recognition”, IEEE Transaction on Image Processing, 19(4): 1087-1096.

[11] Gross, J., Shi, J., and Cohn, J., (2001). “Quo Vadis Face Recognition?”, CMU-RI-TR-01-17, Robotics Institute Carnegie Mellon University, Pittsburg, Pennsylvania 15213.

[12] Tarres, F. and Rama, A., (2005). “A Novel Method for Face Recognition under Partial Occlusion or Facial Expression Variations”, 47th International Symposium ELMAR-2005, Multimedia Systems and Applications, Zadar, Croatia.

[13] Kim, J., Choi, J., Yi, J. and Turk, M. (2005). “Effective Representation Using ICA for Face Recognition Robust to Local Distortion and Partial Occlusion”, IEEE Transactions on Pattern Analysis and Machine Intelligence, 27(12): 1977-1981.

[14] Wright, J., Yang, A.Y., Ganesh, A., Sastry, S.S. and Ma, Y., (2009). "Robust Face Recognition via Sparse Representation", IEEE Transaction on Pattern Analysis and Machine Intelligence (PAMI), 31(2): 210-227.

[15] Arandjelovic, O. and Cipolla, R., (2007). “A Manifold Approach to Face Recognition from Low Quality Video Across Illumination and Pose using Implicit Super-Resolution”, IEEE International Conference on Computer Vision, Rio de Janeiro, Brazil.

[16] Baraniuk, R., (2007). "Compressive Sensing", Lecture Notes in IEEE Signal Processing Magazine, 24.

[17] Candes, E.J. and Wakin, M.B., (2008). "An Introduction to Compressive Sampling", IEEE Signal Processing Magazine, 21-30.

[18] Wakin, M.B., Laska, J.N., Duarte, M.F., Baron, D., Sarvotan, S., Tkhar, D., Kelly, K.F. and Baraniuk, R.G., (2006). "An Architecture for Compressive Imaging", International Conference in Image Processing (ICIP), 8-11 October, Atlanta, GA, USA.

[19] Battini Sönmez, E., Sankur, B. and Albayrak, S., (2011). “Face Classification via Sparse Approximation”, Biometrics and Identity Management Conference (BioID), 8-10 March, Brandenburg University of Applied Sciences, Brandenburg, Germany.

[20] Battini Sönmez, E., Sankur, B. and Albayrak, S., (2011). “Seyrek Yaklaşımlar ile Yüz Sınıflama ” Sinyal İşleme ve Uygulamaları Kurultayı (SIU), 20-22 Nisan 2011, Hacettepe Üniversity Elektrik ve Elektronik Mühendisliği Bölümü, Kemer, Antalya.

[21] Chen, S. and Donoho, D., (1994). "Basis Pursuit", 28-th Annual Asilomar Conference on Signals, Systems and Computers Signals, 1: 41-44.

[22] Boufounos, P., Romberg, J. and Baraniuk, R., (2008). “Compressive Sensing Theory and Applications”, IEEE Internation Conference on Acoustics, Speech and Signal Processing, Las Vegas, Nevada, USA.

[23] Bruckstein, A., Donoho, D.L., Elad, M., (2009). “From Sparse Solutions of Systems of Equations to Sparse Modelling of Signals and Images”, Society for Industrial and Applied Mathematics (SIAM) Review, 51(1): 34-81.

[24] Efron, B., Hastie, T., Johnstone, I. and Tibshirani, R., (2004), “Least Angle Regression”, Annals of Statistics, 32: 407-499.

[25] Pati Y.C., Rezaiifar, R. and Krishnaprasad, P.S., (1993), "Orthogonal Matching Pursuit: Recursive Function Approximation with Applications to Wavelet

Decomposition", 27-th Annual Asilomar Conference on Signals, Systems and Computers, (1): 40 - 44.

[26] Yang, A.,Y., Ganesh, A., Zhou, Z., Sastry, S.S. and Ma, Y., (2010). “A Review of Fast l1-Minimization Algorithms for Robust Face Recognition”, International Conference in Image Processing (ICIP), 26-29 September, Hong Kong, China.

[27] Mallat, S.G. and Zhang, Z., (1993), "Matching Pursuit with Time-Frequency Dictionary", IEEE Transactions on Signal Processing, 41(12): 3397-3415.

[28] Needell, D. and Tropp, A., (2008). “COSAMP: Iterative Signal Recovery from Incomplete and Inaccurate Samples”, Applied and Computational Harmonic Analysis, 26(3): 301-321.

[29] Karahanoglu, N., B. and Erdogan, H., (2011). “A* Orthogonal Matching Pursuit: Best-First Search for Compressed Sensing Signal Recovery”, Int. Conference on Acoustics Speech and Signal Processing (ICASSP), 22-27 May, Prague, Czech Republic.

[30] Wright, J. and Ma, Y., (2010). “Dense Error Correction via l1 Minimization”, IEEE Transaction Information Theory, 56(7):3540-3560.

[31] Brunelli, R., (2010). Template Matching Techniques in Computer Vision, Wiley.

[32] Shi, Q., Eriksson A., van den Hengel, A. And Shen C., (2011). “Is Face Recognition Really a Compressive Sensing Problem?”, IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR), 21-23 June, Colorado Springs, USA.

[33] Lee, K.C., Ho, J., and Kriegman, D., (2005). “Acquiring Linear Subspaces for Face Recognition under Variable Lighting”, IEEE Transaction on Pattern Analysis and Machine Intelligence (PAMI), 27(5): 684-698.

[34] Georghiades, A., Belhumeur, P., and Kriegman, D., (2001). “From Few too Many: Illumination Cone Models for Face Recognition under Variable Lighting and Pose”, IEEE Transaction on Pattern Analysis and Machine Intelligence (PAMI), 23(6): 643-660.

[35] Belhumeur, P. N., Hespanha, J. P. and Kriegman D.J., (1997). "Eigenfaces vs Fisherfaces: Recognition using Class Specific Linear Projection", IEEE Trans. on Pattern analysis and Machine Intelligence, 19(7): 711-720.

[36] Turk, M.A and Pentland, A.P., (1991). "Dimensionality Reduction for Face Recognition", Proc. IEEE Conference on Computer Vision and Pattern Recognition, Maui, Hawaii, 1991, 586-591.

[37] Kanade, T., Cohn, J.F., and Tian, Y., (2000). “Comprehensive Database for Facial Expression Analysis” Proceedings of the Fourth IEEE International Conference on Automatic Face and Gesture Recognition, Grenoble, France, 46–53, 2000.

[38] Lucey, P., Cohn, J.F., Kanade, T., Saragih, J., Ambadar, Z., and Matthews, I., (2010). “The extended Cohn-Kanade Dataset (ck+): A Complete Dataset for Action Unit and Emotion Specified Expression”, Proceedings of IEEE workshop on CVPR for Human Communicative Behaviour Analysis.

[39] Struc, V., (2011). “Performance Evaluation of Photometric Normalization Techniques for Illumination Invariant Face Recognition”, Internal Report: LUKS

[40] Savran, A. and Sankur, B. (2009). “Automatic Detection of Facial Actions from 3D Data”, ICCV’09: Workshop on Human Computer Interaction, September-October, Kyoto.

[41] Ekman, P., Friesen, W., and Hager, J.C., (2002). Facial Action Coding System, Weidenfeld and Nicolson, second edition.

[42] Mehrabian, A., and Ferris, S.R., (1967). “Inference of Attitude from Nonverbal Communication in two Channels”, Journal of Counseling Psychology, 31(3):248-252.

[43] Cotter., S., (2010). “Weighted Voting of Sparse Representation Classifiers for Facial Expression Recognition” IEEE CVPR Workshop on Human Communicative Behavior Analysis, 1164–1168.

[44] Chen, Y., Do, T.T. and Tran, T.D., (2010). “Robust Face Recognition Using Locally Adaptive Sparse Representation”, International Conference in Image Processing (ICIP), 26-29 September, Hong Kong, China.

[45] Davenport, M.A., Duarte, M.F., Wakin, M.B., Laska, J.N., Takhar, D., Kelly, K.F. and Baraniuk, R.G., (2007). "The Smashed Filter for Compressive Classification and Target Recognition", Proc. SPIE Computational Imaging V, San Jose, California.

CURRICULUM VITAE

PERSONAL INFORMATION

Name Surname	: Elena BATTINI SÖNMEZ
Date of Birth and Place	: 1967, Italy
Foreign Languages	: English, Turkish, French
E-mail	: elena@cs.bilgi.edu.tr

EDUCATION

Degree	Given	School/University	Year of Graduation
Master	University of Newcastle upon Tyne, UK		1995
Degree	University of Pisa,	Italy	1993
High School	Pacinotti, La Spezia,	Italy	1986

WORKING EXPERIENCES

Year	Firm/Company	Role
2003-today	Istanbul Bilgi University	Instructor
2003-2004	Yeditepe University	Instructor
1998-2002	Italian Chamber of Trading	Vice-director
1997-1998	Yeditepe University	Instructor

ACADEMIC PAPERS

Conference Papers

1. "Biometric Face Recognition Using Random faces" (original title: Rasgele Vektörlerle Biyometrik Yüz Tanıma), Elena Battini Sönmez, Bülent Sankur, Songül Albayrak, Signal Processing and Communications Applications Conference (SIU2010), 21-23 April 2010, Dicle University, Diyarbakır, TR
2. "Face Classification via Sparse Approximation", Elena Battini Sönmez, Bülent Sankur, Songül Albayrak, Biometrics and Identity Management Conference (BioID2011), 8-10 March 2011, Brandenburg, University of Applied Sciences, Brandenburg, Germany
3. "Face Classification via Sparse Approximations" (original title: Seyrek Yaklaşımlar ile Yüz Sınıflama), Elena Battini Sönmez, Bülent Sankur, Songül Albayrak, Signal Processing and Communications Applications Conference (SIU2011), 20-22 April 2011, Hacettepe, University, Kemer, Antalya, TR

SCHOLARSHIP AND ARCHIEVEMENTS

1. The MSc was supported by a scholarship made available by the University of Pisa. Only one such scholarship per year is awarded.
2. The achievement of the First Class Honour degree was supported by the constant and well organized work done during the university years.