A Memory Based Face Recognition Method

Alex Pappachen James
B. Tech. (Hons), M. Tech.

Griffith School of Engineering
Science, Environment, Engineering and Technology
Griffith University

Submitted in fulfilment of the requirements of the degree of
Doctor of Philosophy

November 2008
Statement of originality

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

Alex. P. James
List of Publications

The following is the paper published during the work done on this thesis, the text of which is included in part in the thesis:

Acknowledgement

This thesis is the account of two years of devoted and rigorous work in the field of pattern recognition at the Queensland Microtechnology Facility in Nathan campus, at the Griffith School of Engineering, Griffith University.

A few lines are too short to make a complete account of my deep appreciation for my supervisors Prof. Sima Dimitrijev and Prof. H. Barry Harrison. It all started two years ago, when they were both my research higher degree supervisors. I wish to thank Prof. Sima Dimitrijev for his unflattering trust and constant encouragements which have been essential to the success of the research in the last two years. I want to extend my appreciation to both of them for their patience, for their full devotion to others and for their great understanding, and to Prof. Sima Dimitrijev for emphasising on the importance of details and most importantly making me slow down with writing, and perhaps also on putting focus on channeled thinking.

Prof. Kuldip K. Paliwal is a distinguished specialist in Speech and Pattern Recognition, whom I first met during one of his visits to the Queensland Microtechnology Facility, and with whom I also had the opportunity to discuss and write a research paper. I wish to thank him for honouring me for the various interactions and friendly discussions.

Prof. Yongsheng Gao is a well regarded expert in the area of face recognition. Prof. Yongsheng Gao was the first person to put forward some of the challenging problems in face recognition, especially by setting a challenge to attain high recognition performance with single gallery images. Although my meetings was rather brief with Prof. Yongsheng Gao, majority of this questions resulted in a fruitful thesis work. I am grateful for that, as well as for his acceptance of my confirmation report, in spite of his very busy schedule.

I would like to thank Prof. A. M. Martinez of CBCL, Ohio State University for providing quick replies and with providing the needed access and support to
the AR face database without which the initial testing and development could have delayed.

During my stay with the Queensland Microtechnology Facility, Griffith University, I have formed various friends who’s interactions was always helpful in keeping the system live and running effectively. I am grateful to Dr. Frederick Kong for helping me settle down in Brisbane in the initial stages and for being a great friend throughout the duration. Neeli M. Rao and Brijesh Patel for arranging regular get-togethers at weekends. And finally to Mitar Milacic for those technical and non-technical discussions.

I would also like to thank the following people, who have provided inspiration and advices for my scientific pursuits over the years, Prof. E.C.G. Sudarshan (Uni. of Texas), Prof. Ashok Rao (IISc), Dr. Sunil Kumar (Texas Inst.), Dr. Ravikishore (HCL Tech.), Dr. M. R. Baiju (Kerala Uni.), Ajayan K. R (IISc), Sajeev G. P. (GEC) and Prof. Achuthsankar S. Nair (Kerala Uni.).

The best of my acknowledgements and love to my parents James and Nirmala, and to my brother Philip. A special thanks to Snoy Francis for being a great friend. Last, but of the most, I would like to acknowledge, my wife Dhanya for her time, patience and never-ending support and love, which helped me tremendously throughout the duration.

I would also like to thank the members of the examination panel including the chair person and external examiners for accepting the thesis for review in spite of their busy schedules and commitments.

Finally, I would like to acknowledge all the great minds, who have contributed to the wealth of knowledge and understanding through various research publications that enabled for an effective and timely research during the preparation of this thesis.
# Contents

List of Figures .......................................................... vi
List of Tables ............................................................. ix
List of Symbols ............................................................ x
List of Abbreviations ....................................................... xiii

## Abstract

1 Introduction ............................................................ 4

1.1 Aim and Scope ....................................................... 4
1.2 Background .......................................................... 5
1.3 Overview of the Approach ............................................ 7

2 Review of Related Work ................................................ 10

2.1 Introduction .......................................................... 10
2.2 Spatial Intensity Changes ............................................. 11
2.3 Similarity Measure ..................................................... 12
2.4 Making Decisions ...................................................... 13
2.5 Early Vision and Face Recognition ................................... 14
2.6 Implementation Issues of Face Recognition ......................... 16

2.6.1 Natural Variability ................................................ 17

2.6.1.1 Illumination .................................................... 17
2.6.1.2 Expression ...................................................... 17
2.6.1.3 Occlusion ...................................................... 18
2.6.1.4 Pose ................................................. 18
2.6.1.5 Localization Error .......................... 18
2.6.1.6 Aging ............................................. 19
2.6.2 Limitation in Training Data .................. 19
2.7 Summary ............................................. 20

3 Local Binary Decisions Based Face Recognition 21
  3.1 Introduction ....................................... 21
  3.2 Local Binary Decisions .......................... 22
  3.3 Spatial Intensity Changes ....................... 23
  3.4 Local Binary Decisions on Similarity Algorithm .... 26
    3.4.1 Feature Extraction ............................ 27
      3.4.1.1 Spatial Change Features .................. 27
      3.4.1.2 Normalized Spatial Change Features ....... 28
    3.4.2 Local Binary Decisions on Similarity Classifier ...... 28
      3.4.2.1 Normalized Similarity Measure .......... 28
      3.4.2.2 Local Binary Decisions .................. 29
      3.4.2.3 Similarity Score and Ranking .......... 30
      3.4.2.4 Localization Error Compensation ....... 30
  3.5 Experimental Details ............................ 31
  3.6 Results and Discussion ........................ 32
  3.7 Conclusions ..................................... 36

4 Analysis of LBDS Algorithm 37
  4.1 Introduction ..................................... 37
  4.2 Experimental Analysis of the Algorithm .......... 38
    4.2.1 Database ...................................... 38
    4.2.2 Analysis and Discussion .................... 39
      4.2.2.1 Effect of Spatial Intensity Change Used as Features 39
## 4.2.2 Normalization

- **4.2.2.2 Normalization** ........................................... 43
- **4.2.2.3 Effect of Mean Normalization and Study of Alternative Normalization** ................. 45
- **4.2.2.4 Effect of Similarity Measure Normalization and Study of Alternative Normalization** ...... 49
- **4.2.2.5 Effect of Local Binary Decisions and Threshold** .................................. 51
- **4.2.2.6 Effect of Resolution** ..................................... 56
- **4.2.2.7 Effect of Color** ........................................ 56
- **4.2.2.8 Effect of Localization** .................................... 58

## 4.2.3 LBDS algorithm

- **4.2.3.1 Feature Vectors** ...................................... 59
- **4.2.3.2 Determination of Similarity** ........................................... 60
- **4.2.3.3 Use of Color Images** ...................................... 60
- **4.2.3.4 Compensating Localization Errors** ........................................... 60
- **4.2.3.5 Parameter Settings** ...................................... 61

## 4.3 Comparison with Other Methods

- **4.3.1 Databases** ........................................... 61
  - **4.3.1.1 AR Database** ........................................ 61
  - **4.3.1.2 Gray FERET Database** ...................................... 62
- **4.3.2 Other Algorithms and Their Settings** ........................................... 62
  - **4.3.2.1 PCA** ........................................ 62
  - **4.3.2.2 PC²A** ........................................ 62
  - **4.3.2.3 SVD PCA** ........................................ 63
  - **4.3.2.4 Other Methods** ...................................... 63
- **4.3.3 Computational Complexity** ...................................... 63
- **4.3.4 Experiments on AR** ...................................... 64
  - **4.3.4.1 Faces with Occlusion** ...................................... 66
  - **4.3.4.2 Faces with Illumination** ...................................... 67
4.3.4.3 Faces with Expressions .......................... 67
4.3.4.4 Comparison with Other Algorithms ............... 67
4.3.5 Experiments on FERET ............................. 68
  4.3.5.1 Faces with Expression .......................... 69
  4.3.5.2 Faces with Illumination ........................ 69
  4.3.5.3 Faces with Aging ............................... 70
4.4 Conclusions ........................................... 70

5 Enhanced LBDS Algorithm ............................... 73
  5.1 Introduction ......................................... 73
  5.2 ELBDS Algorithm ..................................... 74
    5.2.1 Feature Extractor .................................. 76
      5.2.1.1 Spatial Filtering for Information Extraction . . 76
      5.2.1.2 Spatial Change Features ........................ 77
      5.2.1.3 Local Mean Normalization ....................... 79
    5.2.2 ELBDS Classifier .................................. 79
      5.2.2.1 Normalized Similarity and Binary Decisions ..... 79
      5.2.2.2 Global Similarity and Ranking ................ 81
      5.2.2.3 Use of Color Images ............................ 81
      5.2.2.4 Localization Error Compensation ............... 82
      5.2.2.5 Ranking and Best Match ....................... 83
  5.3 Analysis of Spatial Filtering and Classifier .......... 83
    5.3.1 Database and Setup ................................ 83
    5.3.2 Preprocessing Spatial Filters ..................... 84
    5.3.3 Local Binary Decisions on Similarity Averages .... 87
  5.4 Experimental Results and Comparisons ................ 88
    5.4.1 Experiments with AR Database ..................... 88
    5.4.2 Experiments with other Databases .................. 91
      5.4.2.1 FERET Database ............................... 91
CONTENTS

5.4.2.2 YALE Database ........................................... 92
5.4.2.3 Extended Yale Database ................................. 93
5.4.2.4 CALTECH Database ................................. 94
5.5 Conclusions ................................................. 95

6 ELBDS Algorithm in Exampler Based Face Recognition 98

6.1 Introduction ................................................. 98
6.2 Face Recognition Method ................................. 99
  6.2.1 Feature Extraction ..................................... 99
  6.2.2 Classification ........................................... 100
  6.2.3 Compensation of Localization Errors ................. 102
6.3 Experimental Results ...................................... 102
  6.3.1 Equal Number of Training Samples Per Person ....... 102
  6.3.2 Unequal Number of Training Samples Per Person .... 106
  6.3.3 Comparison with Other Methods ....................... 107
    6.3.3.1 AR Database ...................................... 111
    6.3.3.2 ORL, FERET and YALE Databases ................. 111
    6.3.3.3 EYALE Database ................................. 111
  6.3.4 Application of Examplers in Variability Detection ... 113
6.4 Conclusions ................................................. 113

7 Conclusions 115

7.1 Summary of Contributions ................................. 115
  7.1.1 List of Contribution .................................. 115
7.2 Future Directions ........................................... 118
List of Figures

3.1 Block diagram illustrating the various steps in the baseline algorithm. .......................... 24
3.2 An illustration of various steps in the baseline algorithm. .............................. 25
3.3 The dependence of the overall recognition accuracy on the block size used to make the local binary decisions. .......................... 35

4.1 The illustration shows the images of a person in the AR database [1,2] and its organization for the single training sample per person problem depicted in this article. .......................... 38
4.2 Graphical illustrations showing the overall influence of using spatial change features. .......................... 42
4.3 A graphical illustration showing the recognition performance of the LBDS algorithm under the variation of spatial change features filter window size at various image resolutions. .......................... 44
4.4 Graphical illustration showing improved performance of local mean normalization compared to global mean normalization. .......................... 46
4.5 Graphical illustration showing the effect of local mean normalization and similarity measure normalization on the performance of the LBDS algorithm. .......................... 47
4.6 Graphical illustration showing the effect of global mean normalization and similarity measure normalization on the performance of the LBDS algorithm. .......................... 48
<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>4.7</td>
<td>Graphical illustration showing a comparison of normalized similarity measure with a direct similarity measure.</td>
</tr>
<tr>
<td>4.8</td>
<td>Graphical illustration showing the effect of local binary decisions.</td>
</tr>
<tr>
<td>4.9</td>
<td>Graphical illustration showing the stability of the threshold against various normalized similarity measures.</td>
</tr>
<tr>
<td>4.10</td>
<td>Graphical illustration showing the recognition performance of the LBDS algorithm with variation in resolution of the normalized similarity measure $\delta$ under comparison.</td>
</tr>
<tr>
<td>5.1</td>
<td>The figure illustrates the two basic blocks of the ELBDS algorithm using face images from AR database [1, 2].</td>
</tr>
<tr>
<td>5.2</td>
<td>Graphical illustration showing the dependence of recognition performance with the use of number of ELBDS preprocessing spatial filters when using color and gray versions of images in AR database.</td>
</tr>
<tr>
<td>5.3</td>
<td>The azimuth and elevation angles for the 64 illumination conditions in EYALE database.</td>
</tr>
<tr>
<td>6.1</td>
<td>Illustration of preprocessing spatial filters applied to a face image in AR database.</td>
</tr>
<tr>
<td>6.2</td>
<td>Illustration on the organization of AR database [1,2] used for testing.</td>
</tr>
<tr>
<td>6.3</td>
<td>Graphical illustration showing the recognition performance of the proposed method with variation in dimensionality of feature vectors.</td>
</tr>
<tr>
<td>6.4</td>
<td>Graphical illustration showing the time it takes for the classifier for a comparison of a test with all images in the gallery at different feature vector dimensions.</td>
</tr>
<tr>
<td>6.5</td>
<td>A graphical illustration showing the effect of using multiple training samples.</td>
</tr>
</tbody>
</table>
6.6 Graphical illustration showing the recognition performance of the proposed method with variation in dimensionality of feature vectors when the number of training samples used for creating the gallery are randomly selected and are not same for all the persons. 109
List of Tables

3.1 Normalized similarity measures for the classifier . . . . . . . . . 22
3.2 Recognition performance of the LBDS algorithm (Single training sample per person problem) . . . . . . . . . . . . . . . . . . 33
3.3 Summary of the results on different databases . . . . . . . . . . 34
4.1 Effect of global mean normalization and feature type . . . . . 41
4.2 Effect of Global Mean Normalization of Features and Similarity Measure Normalization . . . . . . . . . . . . . . . . . . . . . 44
4.3 Effect of Local Mean Normalization and Distance Normalization 50
4.4 Direct and Normalized Similarity Measures . . . . . . . . . . . 50
4.5 Effect of color on single training samples per person scheme . . 57
4.6 Effect of Localization Error Compensation . . . . . . . . . . . 59
4.7 Parameter Set for Final Algorithm . . . . . . . . . . . . . . . . 61
4.8 Comparison of Complexity . . . . . . . . . . . . . . . . . . . . 63
4.9 Performance of the LBDS Algorithm with a Single Training Sample per Person . . . . . . . . . . . . . . . . . . . . . . . . . . 65
4.10 Comparison of the LBDS Algorithm with Other Algorithms (Single Training Sample per Person Problem) . . . . . . . . . . 66
4.11 Recognition Results on FERET Database . . . . . . . . . . . . 70
5.1 Various Spatial filter operations used in the preprocessing stage of ELBDS algorithm . . . . . . . . . . . . . . . . . . . . . . . . 78
### LIST OF TABLES

<table>
<thead>
<tr>
<th>Table</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>5.2</td>
<td>Spatial filtering window weights of various methods used for extracting various spatial information from an image</td>
<td>80</td>
</tr>
<tr>
<td>5.3</td>
<td>Comparison of information extraction using ELBDS algorithm</td>
<td>84</td>
</tr>
<tr>
<td>5.4</td>
<td>Overall Recognition performance on AR database</td>
<td>88</td>
</tr>
<tr>
<td>5.5</td>
<td>Comparison of the ELBDS algorithm with other algorithms using single gallery image person in the AR database.</td>
<td>90</td>
</tr>
<tr>
<td>5.6</td>
<td>Overall top rank recognition performance of ELBDS algorithm with other databases for single image per person problem</td>
<td>92</td>
</tr>
<tr>
<td>5.7</td>
<td>Comparison of top rank recognition performance of ELBDS algorithm with other algorithms on FERET database</td>
<td>93</td>
</tr>
<tr>
<td>5.8</td>
<td>Recognition performance of ELBDS algorithm for different subsets of EYALE database</td>
<td>95</td>
</tr>
<tr>
<td>6.1</td>
<td>Recognition performance across different databases on the proposed method when the number of training samples in the gallery is fixed for every person.</td>
<td>107</td>
</tr>
<tr>
<td>6.2</td>
<td>The comparison of recognition performance of the proposed ELBDS method for multiple training samples person face recognition problem with other algorithms using AR database.</td>
<td>110</td>
</tr>
<tr>
<td>6.3</td>
<td>The comparison of recognition performance of the proposed ELBDS method for multiple training samples person face recognition problem with other algorithms.</td>
<td>112</td>
</tr>
<tr>
<td>6.4</td>
<td>The comparison of recognition performance of the proposed ELBDS method for multiple training samples person face recognition problem with other algorithms in EYALE database.</td>
<td>112</td>
</tr>
<tr>
<td>6.5</td>
<td>Recognition accuracy of the proposed ELBDS algorithm in the detection of natural variability in the face images.</td>
<td>114</td>
</tr>
</tbody>
</table>
List of Symbols

\( k \)  The index of an image in the gallery of size \( K \)
\( t1 \)  Index of generated synthetic test image
\( g1 \)  Index of generated synthetic gallery image
\( D \)  Size of input raw image
\( N \)  Maximum number of columns in input raw image
\( M \)  Maximum number of rows in input raw image
\( n \)  Maximum number of column in local standard deviation window
\( m \)  Maximum number of rows in local standard deviation window
\( p \)  Index of the texture filter applied
\( P \)  Maximum number of texture filters used
\( c \)  Index of sample for the \( k^{th} \) person
\( v \)  Number of vertical perturbations
\( h \)  Number of horizontal perturbations
\( s \)  Total number of training samples
\( E \)  Decision block size (in pixels)
\( o \)  Number of overlapping decision partitioned regions of size \( E \)
\( f \)  Reduced dimensionality of features, \( f \ll D \)
\( I \)  Input raw image
\( I(i,j) \)  Pixel in input raw image at spatial location \((i,j)\)
\( \overline{I(i,j)} \)  Local mean of a pixel \( I \)
\( I_g^{(k)} \)  \( k^{th} \) gallery image
\begin{itemize}
  \item $I_t$ Any test image
  \item $I_g^{t1}$ Any person with $t1^{th}$ synthetic gallery image
  \item $I_g^{d1}$ $d^{th}$ person with $g1^{th}$ synthetic gallery image
  \item $I_g^{(k,c)}$ The $c^{th}$ sample of $k^{th}$ gallery image
  \item $\sigma$ Local standard deviation vector
  \item $\overline{\sigma}$ Global mean of $\sigma$
  \item $\sigma^{(p,*)}$ Local mean of $\sigma^{(p,*)}$
  \item $\sigma^{(p,*)}$ Spatial change detection on $Y^{(p,*)}$
  \item $x$ Normalized spatial change feature vector
  \item $x_g$ $x$ of gallery image
  \item $x_t$ $x$ of test image
  \item $x_g^{t1}$ $x$ from any person with $t1^{th}$ synthetic gallery image
  \item $x_g^{d1}$ $x$ from $d^{th}$ person with $g1^{th}$ synthetic gallery image
  \item $x^{(p,*)}$ $x$ calculated from $\sigma^{(p,*)}$
  \item $x^{(*)}_r$ $x$ from red channel of color image
  \item $x^{(*)}_g$ $x$ from green channel of color image
  \item $x^{(*)}_b$ $x$ from blue channel of color image
  \item $x_g^{(p,k)}$ $x$ from $p^{th}$ texture filter and $k^{th}$ gallery image
  \item $x_t^{(p)}$ $x$ from $p^{th}$ texture filter and any test image
  \item $x_g^{(p,k,c)}$ $x$ formed by applying $p^{th}$ texture filter on $c^{th}$ sample image of $k^{th}$ person in the gallery
  \item $\theta$ Global threshold
  \item $\gamma$ Similarity measure normalization factor
  \item $\delta$ Normalized similarity measure
  \item $\delta^{(k)}$ $\delta$ as a result of comparison between $x_g^{(k)}$ and $x_t$
  \item $\delta^{(p,k)}$ $\delta$ for the comparison of $k^{th}$ gallery with any test image for features from $p^{th}$ filter
\end{itemize}
\( \delta^{(p,k)} \) Average of \( \delta^{(p,k)} \) from filters \( p = 1 \ldots P \)

\( \delta^{(p,k,c)} \) \( \delta \) formed from comparison of \( x_g^{(p,k,c)} \) and \( x_t^{(p)} \)

\( \hat{\delta}^{(k,c)} \) Average of \( \delta^{(p,k,c)} \) from filters \( p = 1 \ldots P \)

\( w \) Local standard deviation window

\( w_f^{(p)} \) Texture filter window with index \( p \)

\( B_0 \) Local binary decisions vector formed from any \( \delta \)

\( B_0^{(k)} \) \( B_0 \) formed from \( \hat{\delta}^{(k)} \)

\( B^{(k)} \) Local binary decisions vector formed from \( \hat{\delta}^{(k)} \)

\( B^{(k,c)} \) Local binary decisions vector formed from \( \hat{\delta}^{(k,c)} \)

\( S_g \) Global similarity score

\( S_g^{(k)} \) \( S_g \) for the comparison between \( x_g^{(k)} \) and \( x_t \)

\( S_g^{(1)} \) \( S_g \) for the comparison between \( x_g^{(k)} \) and \( x_t^{(1)} \)

\( S_g^{(q)} \) \( S_g \) for the comparison between \( x_g^{(k)q} \) and \( x_t \)

\( Y^{(p,*)} \) The result of \( p^{th} \) texture filtering on an input image

\( Y^{(p)}_t \) The result of \( p^{th} \) texture filtering on any input test image

\( Y^{(p,*)}_g \) The result of \( p^{th} \) texture filtering on an input gallery image

\( Y^{(p,d)}_g \) The result of \( p^{th} \) texture filtering on an input gallery image of \( k^{th} \) person

\( Y^{(*)} \) Local mean of \( Y^{(p,*)} \)
### List of Abbreviations

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>PCA</td>
<td>Principle Component Analysis</td>
</tr>
<tr>
<td>ICA</td>
<td>Independent Component Analysis</td>
</tr>
<tr>
<td>LDA</td>
<td>Linear Discriminant Analysis</td>
</tr>
<tr>
<td>KPCA</td>
<td>Kernel Principal Component Analysis</td>
</tr>
<tr>
<td>SVDPCA</td>
<td>Singular value decomposition and principal component analysis</td>
</tr>
<tr>
<td>ISOMAP</td>
<td>Isometric Feature Mapping</td>
</tr>
<tr>
<td>LLE</td>
<td>Locally Linear Embedding</td>
</tr>
<tr>
<td>EBGM</td>
<td>Elastic Bunch Graph Matching</td>
</tr>
<tr>
<td>AAM</td>
<td>Active Appearance Model</td>
</tr>
<tr>
<td>SIS</td>
<td>Single Image Subspace</td>
</tr>
<tr>
<td>LBP</td>
<td>Local Binary Patterns</td>
</tr>
<tr>
<td>LBDS</td>
<td>Local Binary Decisions on Similarity</td>
</tr>
<tr>
<td>ELBDS</td>
<td>Enhanced Local Binary Decisions on Similarity</td>
</tr>
<tr>
<td>DVD</td>
<td>Digital Versatile Disc</td>
</tr>
<tr>
<td>FERET</td>
<td>Facial Recognition Technology</td>
</tr>
</tbody>
</table>
Abstract

The human brain exhibits robustness against natural variability occurring in face images, yet the commonly attempted algorithms for face recognition are not modular and do not apply the principle of binary decisions made by the firing of neurons. This thesis presents a memory based face recognition method based on the concepts of local binary decisions and spatial change features. Local binary decisions are inspired from the binary conversions done by firing of neurons while spatial change features are inspired from the retinal processing of the human visual system. Applying these principles and by using the principle of modularity in a hierarchical manner, a class of memory based face recognition algorithms is formed.

These algorithms when applied to difficult testing conditions show high recognition performance. This high recognition performance is enabled by (1) local binary decisions and (2) spatial change detection. The baseline algorithm formed by using these two concepts is called local binary decisions on similarity (LBDS) algorithm. An analysis is performed using the LBDS algorithm to optimize the parameters, and to study the relative effect of spatial change features, local binary decisions, normalization of features, normalization of similarity measure, use of color, localization error compensation and resolution on recognition performance. From the insights gained through the analysis, the LBDS algorithm is further improved by incorporating various preprocessing spatial filter operations.
to extract more spatial information. The inclusion of preprocessing step helps to achieve even higher recognition performance and robustness to difficult tasks. This improved algorithm is called *enhanced local binary decisions on similarity* (ELBDS) algorithm. The ELBDS algorithm is further used to incorporate the multiple training images per person in the gallery, and is called an exemplar based face recognition method.

The following is the overall recognition performance when using single gallery image per person: 97% on AR, 100% on YALE, 97% on EYALE, 97% on CALTECH, 98% on FERET(FaFb), 94% on FERET(FaFc), 74% on FERET(FaDup1) and 76% on FERET(FaDup2). When using multiple training samples per person, following recognition accuracies are achieved, 99.0% on AR, 99.5% on FERET, 99.5% on ORL, 99.3% on EYALE, 100.0% on YALE and 100.0% on CALTECH face databases.

**Summary of Contribution**

(1) Local binary decisions is identified as an important concept that is required for recognition of faces under difficult conditions.

(2) Spatial intensity changes in the face is identified as the visual cue for face recognition.

(3) A performance analysis showing the relative importance of local binary decisions, spatial change as features, color images, compensation techniques to handle localization errors, and various normalization techniques for feature representation and classification stages is performed.

(4) Preprocessing spatial filter operations to extract spatial information is identified as an important preprocessing step to increase robustness against natural variability.
The use of examplers as a method to incorporate multiple training samples per person in the gallery set is proposed.

A successful implementation of the principle of modularity to increase system complexity and/or stability is presented through the formation of a class of memory based face recognition algorithms.
Chapter 1

Introduction

1.1 Aim and Scope

The focus of this thesis is to present a simple and robust face recognition method. In formulating such a recognition method, two important aspects are required: (1) extract visual features that contain the uniqueness of a face called the identity information, and (2) classify the face based on its identity. In order to extract the identity information for comparison, it is important to understand the visual cues that enable the recognition of faces. For the classification, it is important to make the process of decision making simple and effective. The main aims of this thesis are summarized as follows:

✓ To investigate and design a method for extracting facial features.

✓ To design a simple classifier for comparing the extracted facial features.

✓ To design an easy-to-implement software or hardware algorithm for face recognition utilising the designed feature extractor and classifier.

✓ To achieve robust performance against various natural variability.
1. Introduction

To apply the principle of modularity in a hierarchical manner to form a robust face recognition system.

1.2 Background

Face recognition is a special topic in visual information processing that has grown to be of tremendous interest to pattern recognition researchers for the past couple of decade [3–7]. The reason for this interest has been due to the wide range of its practical applications in person identification, photo searching, surveillance, human computer interactions and multimedia technologies.

In humans, the process of face recognition is a dedicated and unique task which is enabled by the use of biological memory. Further for face processing [8], memory is used for the comparison of the identity information from a test face to that of the stored identity information corresponding to various gallery faces. The best match among all the resulting comparisons between a test face and a set of gallery faces results in the identification of that test face with some accuracy. The accuracy with which the recognition occur depends on, (1) the amount of identity information available from a face, and/or (2) how well the learning mechanism is able to extract the available identity information. It can be seen that a mechanism to find identity information and a method of comparison are required to find the best match of an unknown test face to a known representation of faces stored in memory. Translating these ideas of human face recognition to machines results in two essential blocks: (1) feature extraction, and (2) classification. Feature extraction process aims at the extraction of identity information, while classification tries to find the best discrimination between the comparing faces.

However, in the case of machines, recognition of faces in images with various natural variability is a difficult task. Natural variabilities that effects the per-
formance of a face recognition system include, (1) illumination effect on faces, (2) expression of faces, (3) variation in pose, (4) occlusion in face images, (5) aging that deforms the faces with respect to time, and (6) improper localization and alignment. Most image based face recognition methods can be classified into two types: appearance and model based methods [9]. Performance of every face recognition algorithm is effected by one or more of the above mentioned natural variability. The classical linear appearance-based classifiers based on, Principle Component Analysis (PCA) [10–12], Independent Component Analysis (ICA) [13] and Linear Discriminant Analysis (LDA) [14] employ the idea of dimensionality reduction and compression to compensate for the effects of natural variability. In the past, to increase the robustness of appearance-based methods against natural variability, nonlinear (manifold) analysis methods like Kernel Principal Component Analysis (KPCA) [15,16], Isometric Feature Mapping (ISOMAP) [17] and Locally Linear Embedding (LLE) [18] were attempted [9]. On the other hand, model-based methods use feature based comparisons by constructing a face model based on some prior knowledge. The test face is compared by fitting its model parameters on the prototype face. Some of the most widely used methods that fall under this scheme are Elastic Bunch Graph Matching (EBGM) [19] and Active Appearance Model (AAM) [20]. Despite a lot of research, all of the mentioned methods suffer from the effect of natural variabilities and do not show robust performance against difficult conditions.

The recognition performance becomes even worse due to natural variability in face images when the gallery set is limited to single training sample per person [21]. This situation is often called as single gallery image per person problem. Although this is a technically difficult problem to address, it has various biometric applications. By limiting the number of training samples to one, many of the methods that rely on multiple number of training samples per person (for example subspace based methods [22]) for forming models or features fail to
work or perform very poorly. Further, with an increase in natural variability the classic face recognition systems are no longer reliable for use. On the other hand, fewer samples per person results in better computational efficiency and enables faster learning. Some of the very recent and/or well known algorithms that address this specific problem are, (PC)

A \cite{[23]}, Single Image Subspace (SIS) \cite{[24]}, SOMface \cite{[25]}, SVD-PCA \cite{[26]}, EBGM \cite{[19]}, Local Binary Patterns (LBP) \cite{[27]}, LPS \cite{[28]}, Modular FLDA \cite{[29]} etc. These methods use either one or a combination of the following techniques, (1) forming multiple synthetic images from a single gallery image that addresses the effects of pose variation, illumination variation and alignment errors, (2) use of parametric tuning or perturbation on images to enable subspace calculations, and (3) designing better matching schemes and classifiers (eg. probabilistic matching and neural network methods) to increase classifier performance against random variability \cite{[25,28,30–33]}.

Robustness against various difficult conditions is essential for improving the recognition performance of face recognition algorithms. In addition, a simple design with robust performance against difficult tasks can help in the formation of a reliable complex system that can be useful to general visual pattern recognition problems.

1.3 Overview of the Approach

To have a face recognition algorithm that is invariant to all variabilities is a very difficult task. There are many possibilities and ways of handling such variabilities: (1) one can think of dealing with each variability individually and construct a feature vector that is invariant to all the known variabilities, (2) form a robust feature extraction method that is invariant to most of the variability and then improve the system performance by other known compensation techniques, and (3) use multiple images to form the gallery images of a person such that all the
possible variabilities are accounted for.

In humans, it is possible that all of these techniques are employed in a way or other and it becomes logical to try and implement the combination of techniques in the implementation of a face recognition algorithm in machines. In an attempt to emulate this, a biologically inspired face recognition method is developed that is simple and robust; and employs the principle of modularity, local binary decisions on similarity of features under comparison and spatial intensity changes in image as feature. Local binary decisions on similarity methods are formed of two components: (1) a feature extraction and, (2) a classifier based on local binary decisions. Utilizing these concepts various algorithms are presented: (1) Local Binary Decision on Similarity (LBDS) algorithm, (2) Enhanced Local Binary Decision on Similarity (ELBDS) algorithm and (3) an exemplar based face recognition algorithm. Since these algorithms involve a direct storage and comparison using memory, it is logical to call these algorithms in general as based on memory based face recognition method. The following four chapters in this thesis presents the method of using local binary decisions on similarity of features in face recognition:

1. LBDS algorithm (Chapter 3)

In this chapter, local binary decisions on similarity algorithm is presented. The two main blocks, (1) spatial change detection, and (2) local binary decisions based classifier are explained. The use of these two blocks in face recognition is shown through tests on various publicly available face recognition databases. The recognition performance is studied for single gallery image per person face recognition problem.

2. Analysis of LBDS algorithm (Chapter 4)

The individual components in LBDS algorithm are analysed for studying the effects of feature normalization, similarity measure, threshold, use of
1. Introduction

color images and localization error compensation. This analysis helps in a better understanding, enables optimization of parameters, and helps to benchmark the recognition performance of LBDS algorithm in comparison with other well known algorithms.

3. ELBDS algorithm (Chapter 5)

The ELBDS algorithm improves the LBDS algorithm significantly by including a texture based spatial filtering scheme followed by spatial change filtering. Also an average similarity measure is formed to improve the correctness of local binary decisions.

4. Exemplars for face recognition (Chapter 6)

The ELBDS algorithm is further used to study the effect of multiple image per person problem.
Chapter 2

Review of Related Work

2.1 Introduction

Understanding vision in humans at the level of forming a theoretical framework suitable for computational theory, has opened up various disagreements about the goals of cortical processing. The works of David Marr and James Gibson are perhaps the only two major attempts to provide deeper insight.

In majority of Marr’s work [34], he assumed and believed vision in humans to be nothing more that a natural information processing mechanism, that can be modelled in a computer. The various levels for such a task would be: (1) computational model, (2) a specific algorithm for that model, and (3) a physical implementation. It is logical in this method to treat each of these level as independent components and is a way to mimic the biological vision in robots.

Marr attempted to set out a computational theory for vision in a complete holistic approach. He applied the principle of modularity to argue visual processing stages, with every module having a function. Philosophically, this is one of the most elegant approach proposed in the last century that can suit both the paradigms of software and hardware implementations.

Gibson on the other hand had an ecological approach to studying vision. His
view was that vision should be understood as a tool that enables animals to achieve the basic tasks required for life: avoid obstacles, identify food or predators, approach a goal and so on. Although his explanations on brain perception were unclear and seemed very similar to what Marr explained as algorithmic level, there has been a continued interest in the rule-based modeling which advocates knowledge as a prime requirement for visual processing and perception.

Both these approaches have a significant impact in the way in which we understand the visual systems with the aim to form a formidable computational model useful to pattern recognition and/or machine learning. In the following section provides a brief historical perceptive on three major concepts: (1) spatial intensity changes in images, (2) similarity measures for comparison, and (3) decision making using thresholds, which can be considered very important to support the method proposed in this thesis.

## 2.2 Spatial Intensity Changes

Any image that falls on the human eye has varying light intensities depending on the surface reflectance of the object one looks at, and/or strength of the light source, its direction and distance of the viewpoint. However, as known from various computational and physiological experiments, sudden spatio-temporal changes in intensities stimulate the sensory neurons in the eye and is believed to be the main source of information that results in the ability to characterise an object [35–38]. Majority of such studies find a high response to stimuli for sharp changes, or in other words to edges in an image.

Influenced by these understanding and in some cases by logical or observational thinking, many researchers over the last century has proposed various edge filtering schemes. The busyness of the edges obtained from the filtering schemes differ from one method to another. However, in general there will be
large amount of sharper image edges if only a convolution window of few pixels are used (or more the number of pixels the less sharper the gradient for the same image under consideration). This in a way can be due to inter-pixel correlation, the more the number of pixels one considers more is the chance for the detected edge to become distorted. These inter-pixel correlations are also very important for understanding the visual perception of the natural images such as faces, natural scenes and objects [39–43]. The question of forming a right computational model for detecting an edge in an image is very quantitative in nature, and its measure depends on the ability of the brain to measure similarity between two comparing edges and hence objects.

2.3 Similarity Measure

The concept of similarity is of fundamental importance to the perception of patterns that are used for various comparisons. These comparisons are used in recognition of visual patterns, sensory signals and in language processing. In pattern recognition, they find application in the implementation of classifiers. Similarly, in mathematics, they find application in semantics, geometry and taxonomy. It is almost impossible to do a justified review, because of it wide use. However, we will try to address and justify the techniques that are of importance or are influenced into use in the pattern recognition arena.

Distance measure or similarity measure or often known as Multidimensional scaling (MDS) uses the concept of similarity as inversely related to distance. An example to this type of similarity measure is Euclidean distance. In the aim to understand perception of similarity, one of the first form of MDS used data in the form of dissimilarity judgments which were measured on an interval or ratio scale [44]. These algorithms were later generalized into non-metric MDS, which requires ordinal scale data [45, 46]. A later development, which has seen
to generate interest recently, accounted for individual differences in a version of non-metric MDS that assumes people produce different similarity judgments because they differentially weight the various stimulus dimensions [47–50]. An interesting law was proposed by Shepard which he called as a universal law. In this law the distance and perceived similarity are related via an exponential function [51]. But he noted it to fail when the objects are confusable (or very similar). When talking about distance measure, it is a common practice to believe that all the distance measures obey the so called distance axioms: (1) Equal self-similarity, (2) Minimality, (3) Symmetry, and (4) Triangle Inequality. However, such mathematical basis of formulating distance measures has often been questioned and in many cases seen to fail [52]. Further, in response to empirical evidence against the distance axioms, Tversky proposed that perceived similarity is the result of a feature-matching process that differentially weights common and distinct stimulus features. And this in a way advocates the need for feature based similarity comparison. In the context of face recognition, it is evident that the methods with robust features as opposed to raw image intensities perform better on a same measure of similarity. Further, it also points to the assumption that any distance measure can be used for measuring similarity as far as a judgement can be made on it.

2.4 Making Decisions

Judgements or decisions are always based on a threshold or a set of thresholds. One can argue that the decisions can be made by probability measures, however, it will also require a decision boundary or threshold to make any judgement. Neurons can be thought of as the early markers for the origin of decision making and hence intelligence in humans. Much of the studies on thinking processes come from the electrophysiological recordings of single neuron. As noted by
Deadwyler and Hampson, this approach of reconstructing the function of brain is like trying to decipher a video image one pixel at a time while the video image constantly changes [53].

In the biological neuronal systems, modules in the brain cooperates directly or indirectly with others and has various characteristics: (1) the unequal roles of individual components, (2) a presence of excitatory and inhibitory elements, (3) a dependence on the previous history of the system [54], (4) mutual interactions of many kinds due to temporal and spatial convergence and divergence, (5) interactions lasting hundreds of milliseconds and longer, and still maintaining high accuracy [55], (6) multiplication and summation operations [56], (7) parallel and serial processing, and (8) numerous and multiple feedback and feedforward loops with different delays which are able to generate periodic or aperiodic chaotic rhythms [56]. It can be noted that the process of decision making in human brain is complex, yet relatively the structural and functional components that enables this is very simple. Neurons are not digital switches. Rather, they are analog computing devices integrating many inputs and producing an all-or-nothing digital output. In terms of functionality neurons are logic function generators that makes decisions enabled by its firing threshold.

2.5 Early Vision and Face Recognition

Over the years there has been some significant ground covered in the understanding of human visual system. This wealth of knowledge adds a lot of value to computational vision research. Although, one can argue that face recognition can be achieved in a pure theoretical frame of mind by applying mathematical algorithms, it is always better to look at and draw inspiration from human visual system that is known to work and produce robust results.

The discovery of orientation-specific cells in the feline primary visual cortex,
paved way to the understanding of a hierarchical model of visual processing, in which, lines and edges were used to construct more complicated forms [57]. Further, these ideas were seen to show influence in the work of Marr in his book *Vision* [34].

The low-level vision studies has been utilised in the *Malsburg Model* for face recognition. Attempts have been made in the past to model the low-level vision in early visual cortex. In these models, the receptive fields of the cells are modeled as *Gabor functions*, which is useful in the manipulation of frequency, orientation and size. Using these functions to perform recognition, a *Gabor jet* is used as a strategy to extract the various required measurements. They were successfully applied to techniques such as EBGM [19]. Interestingly enough, this method shows similarity values to correlate strongly with human similarity judgements [58]. This example, illustrates that models of human early vision can be made use in face recognition with some success. However, it is worthwhile to mention that not much of face recognition algorithm pay enough attention to the capabilities provided by human face recognition.

Sinha et. al. put forward 19 important results from the understanding gained from human face recognition [59]. This mainly dealt with resolution, color, holistic processing, importance of eyebrows, encoding, illumination influence on generalization etc.

The ability of humans to identify low-resolution face images was studied by Harmon and Julesz on block averaged images of familiar faces, they found high recognition rates even with $16 \times 16$ pixels [60,61]. However, there is not enough evidence that suggest that this is accomplished by low-level vision and is believed to be due to high level thinking associated with the brain. This type of trick that brain does is also evident from the way in which human brain deals with very familiar face. Familiar faces are easy to recognize even with degradation [62,63].

Another, important finding that has a vital role in face recognition is the
feature representation. The features that contributes to recognition of faces was studied by various researchers in the past. The very early psychological experiments on children showed that face fixations occur 30% to the eyes while 60% to that in edges, which indicates to the relative importance of eyes and edges in recognition of faces [64]. However, on the contrary, the studies on high spatial frequencies suggest that sharp edges plays lesser role in recognition than was anticipated [65–67]. This further suggests that low-level edge maps or medium edges in the image also plays important role in the robust recognition of faces [68].

Holistic processing of face is yet another characteristic of human visual system. This process involves an interdependency between features and configurational information. Few work, that attempted to explore this nature are holistic information based learning [69] and facial expression [70]. Holistic processing can be seen as first level processing that is important and unavoidable. However, it is also important to understand that especially when the faces become more similar and known, the local features tends to play a major role. Each feature in the human face tends to show different level of weight in contribution to its identity [71–73].

2.6 Implementation Issues of Face Recognition

The two major problems that has a significant impact on the recognition performance of computational face recognition methods are: (1) natural variability and (2) limitation in training data. In this section, an overview of these problems and the efforts put in to find an invariance to those problems are provided.
2. Review of Related Work

2.6.1 Natural Variability

Natural variability in face images occur due to different kinds of spatio-temporal variation: (1) illumination, (2) expression, (3) occlusion, (4) pose, (5) localization errors and (6) aging.

2.6.1.1 Illumination

In humans, the illumination changes do not effect the way in which we perceive the identity of a face. However, in computational pattern recognition this is often not an easy task. Any change in the illumination changes the intensity values, which in effect demands for a technique that can either compensate for such variations or that can form illumination invariant feature vector. In the case of 2-dimensional face images, methods that try to achieve invariance to illumination are: image gradients [74], Gabor jets [75], the direction of image gradients [76,77], and projections to subspaces derived from linear discriminants [78]. However, an obvious extension to these methods are to use the 3-dimensional information which may be more powerful, as they have the potential to compensate completely for lighting changes, whereas 2-dimensional methods cannot achieve such invariance [76,79,80].

2.6.1.2 Expression

Facial expressions can trick the face recognition systems due to the configurational changes in the location of features especially from eyebrows, eyes and mouth. One approach to achieve expression invariance is by morphing images to become the same shape as the one used for training. However, automatic morphing is not an easy process often due to the lack of texture information in the training data. Another approach is to use optical flow. However, it is difficult to learn the local motions within feature space to determine the expression changes of each face, since different persons express a certain expression in
different ways. Finally, a method of weighting [81], that weights independently those local areas which are less sensitive to expressional changes.

2.6.1.3 Occlusion

Occlusions causes the loss of identity information. The problem of partially occluded faces are addressed by local feature analysis methods. In these techniques the features in the face are divided into different parts and, then, a voting space is used to find the best match [81]. Another, way is to detect occlusion by separating out the facial features needed to be compared. Further, the issue of occlusion is more dominant when the training sample per person in the gallery is limited to one.

2.6.1.4 Pose

The problem caused by pose variations is reduced by the use of two major techniques. One depends on the use of multiple gallery images per person for face recognition and the other makes use of 3-dimensional information of the face for improving accuracy. Multiple gallery images per person are mainly used in appearance-based methods like Eigenfaces and SLAM [82]. An obvious limitation to this method is the lack of robustness when the number of gallery images per person are limited. To reduce this problem, parallel deformation techniques were used to synthesize virtual views from a real view using two dimensional images of rotating prototype faces [83,84]. This concept of virtual view was extended to three dimensional face model for synthesizing virtual views for invariant face recognition [85,86].

2.6.1.5 Localization Error

Localization errors are often considered as a result of improper localization that occurs after face detection process. Anyway, this can be also considered as
a type of natural variability that occur in the case of any machine based face recognition system. The effect of localization is not studied well, as much of these is compensated when multiple gallery images per person are used. However, in the case where single training image per person is used in forming the gallery, the effect of localization error is significant. This issue is addressed in the work of Martinez [2]. A more recent method called single image subspace (SIS) tries to compensate for such errors by using shifted version of the gallery images to generate synthetic images [24].

2.6.1.6 Aging

The effect of aging of faces on face recognition performance in real photo verification task was studied by Ling et. al. [87]. Although their results showed that the aging process increases the difficulty, it does not however surpass the influence of illumination or expression. However, as these latter issues, namely, illumination and expression, are being successfully addressed by incorporating 3-dimensional models, aging process will continue to be a major obstacle for performance improvement [88,89].

2.6.2 Limitation in Training Data

Limitation in training data is one of the most practical issues which was neglected in face recognition algorithm design for many years. However, in recent years, active research has been pursued in achieving high recognition performance with single training image per person for forming the gallery set [21]. The main challenge with this problem is achieving robust performance under varied natural variability. This is considered as one of the most difficult problems in face recognition. Some of the techniques that have been attempted to attack this problem are by synthesizing virtual samples, localizing the single training image, probabilistic matching and neural network methods [25,28,30–33]. More recent
techniques such as SIS [24] have shown to have good recognition performance against different natural variability. However, a more rigorous investigation is required to make face recognition robust with limited training data to be practical to real-time applications. Further, as an obvious task, an improved recognition against such difficult situation would imply a higher recognition performance with more training images per person.

2.7 Summary

Studies on biological vision has revealed various information useful for pattern recognition research. Most approaches in use today are a result of direct or indirect implication from the studies on visual process that occur in humans. However, very few methods try to emulate the biological functions in forming a computational model or method for achieving face recognition task. Three important aspects are discussed with a view to form a background on face recognition task: (1) studies on spatial change as a visual cue, (2) idea of similarity measure for achieving a quantitative formulation of difference between the comparing quantities, and (3) method of decision making process in humans. There are two main reasons that make face recognition difficult: (1) natural variability, and (2) limitation in training data. The coexistence of these two problems makes the present-day face recognition algorithms unreliable and in many cases, causes it not to work. Over, the last decade significant progress has been made to deal with individual natural variabilities, however, some questions still remain open, including overall robustness.
Chapter 3

Local Binary Decisions Based Face Recognition

3.1 Introduction

The fundamental problem that hinders the development of a robust automatic face recognition system is the natural variability that occurs in face images. Major methods that are employed to reduce this problem can be classified into three groups: (1) methods whose so-called gallery set consists of multiple training images per person [90, 91] (2) image preprocessing techniques that aim at feature restoration [92], and (3) use of geometrical transforms to form face models [92]. Even though they show high performance under specific conditions they lack overall robust performance and in many cases have proved to be computationally expensive. As distinct from these computational schemes, the human visual system, which is the best available natural model for face recognition, uses modular approach for classification of faces [93].

This chapter presents a method that implements the concept of local binary decisions to form a modular unit and a modular system for face recognition. This method is applied to formulate a simple algorithm and its robustness verified
3. Local Binary Decisions Based Face Recognition

against various natural variability occurring in face images.

3.2 Local Binary Decisions

The simplest form of making a decision is by the assertion of TRUE or FALSE on a given decision problem. Such decisions are called binary decisions as only two forms of choices exist. Binary decisions (or yes/no decisions) are inevitable in any classification task. In the case of human brain, these decisions are made by the firing of its neurons, which means that they are local binary decisions. Most existing face recognition approaches utilize probabilistic or deterministic formulations to delay the binary decisions until a global level is reached. However, it is well established in digital electronics that local binary decisions can remove noise in a way that is not possible in analog systems, regardless of whether probabilistic or deterministic formulations are used in the analog system. Inspired by these analogies, local binary decisions are applied to every pixel of the images that are being compared.

Table 3.1: Normalized similarity measures for the classifier

<table>
<thead>
<tr>
<th>Type</th>
<th>Equation$^a$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Min-max ratio</td>
<td>$\min[x_g, x_t]/\max[x_g, x_t]$</td>
</tr>
<tr>
<td>Difference</td>
<td>$</td>
</tr>
<tr>
<td>Exponential difference</td>
<td>$e^{-</td>
</tr>
</tbody>
</table>

where $\gamma$ is $\max[x_g, x_t]$ or $[x_g + x_t]/2$ or $\min[x_g, x_t]$

$^a$ The vector multiplications and divisions are element by element operations. It can be also noted that choice of $\gamma$ does not alter the recognition performance.

In addition, the binary decisions necessitate the use of a threshold $\theta$ so that some differences can be ignored as the decisions on similarities are made. The most common method for similarity matching is the use of various distance measures [94]. The similarity measures need to be normalized to enable their use with a global threshold. A normalized comparison between two normalized feature
3. Local Binary Decisions Based Face Recognition

vectors $x_g$ and $x_t$, formed from the gallery and test images, respectively, can be achieved in many ways. The simplest way to form a normalized measure $\delta_{gt}$ is by taking an element by element ratio of the minimum and the maximum of the vectors $x_g$ and $x_t$ (Table 3.1). Other possibilities, also shown in Table 3.1, include normalized difference and exponential difference. Thresholding the similarity measure results in a binary vector representing the local decisions made by the element-wise comparisons. Taking the sum of all the elements in this vector forms the global similarity score for the two variables. Following the comparisons of the test image to all the gallery images, the similarity scores are ranked. This completes the steps of what is commonly referred to as a classifier.

3.3 Spatial Intensity Changes

The raw image vector, which could be the simplest feature vector, is highly sensitive to variation in illumination. This problem can be reduced by using spatial intensity changes as the feature vector. Furthermore, a number of published results suggest that the spatial intensity change is the essential visual cue required for recognition [95]. There are many ways of calculating the spatial intensity changes from a raw image vector. Vectors of spatial intensity changes $\sigma_g$ from gallery image $I_g$ and $\sigma_t$ from test image $I_t$ are calculated by using any of the following standard local spatial filters$^1$: standard deviation, range, and gradient. Further, to obtain the feature vectors $x_g$ and $x_t$, normalization are performed by dividing each element of $\sigma_g$ and $\sigma_t$ by their regional means, $\sigma_g$ and $\sigma_t$, calculated using centered local windows of constant size.

$^1$These filters are available in MATLAB 7 as a part of the image processing library, whose function names are: stdfilt, rangefilt, and gradient.
Normalized Similarity measure calculation

Apply local binary decisions to form local binary decisions vector

Calculate similarity score

Final decision by ranking

Figure 3.1: Block diagram illustrating the various steps in the baseline algorithm. The normalized spatial change features are extracted from the test image and the gallery image using a feature extractor. These normalized features are then compared using the LBDS classifier. The similarity score shows a numerical measure of global similarity between the two images under comparison.
3. Local Binary Decisions Based Face Recognition

Figure 3.2: An illustration of various steps in the baseline algorithm. The images labeled (a), (b), and (c) show the raw images, where (a) and (c) form the gallery images and (b) is a test image, all taken from the AR database [1]. The images labeled (d), (e), and (f) show the output of a feature extraction process described in Section 4.2.2.1, which corresponds to the raw images (a), (b), and (c), respectively. The normalized feature vectors described in Section 3.4.1.2 are shown as the images labeled (g), (h), and (i), and are calculated from (d), (e), and (f), respectively. This is followed by comparisons of test image with gallery images. The normalized similarity measure described in Section 3.4.2.1 when applied for comparing (h) with (g) and (h) with (i) results in images labeled (j) and (k), respectively. Finally, the local binary decisions described in Section 3.4.2.2 when applied on (j) and (k) result in binary vectors labeled (l) and (m), respectively. Clearly, in this example, (b) is a best match to (a) due to more white areas (more similarity decisions) in (l) than in (m).
3.4 Local Binary Decisions on Similarity Algorithm

Most face recognition algorithms use intensity pixel values in raw image $x$ as its features, which is followed by an efficient coding scheme and a classifier for classification. As distinct from these, the Local Binary Decisions on Similarity Algorithm (LBDS) algorithm for face recognition consists of a feature extractor and a classifier. The feature extractor is used to extract features from the raw image to form a feature vector $x$. Vectors thus formed are then used in the classifier for classification.

The following sections, describes a feature extractor that extracts spatial intensity change features [which is commonly referred as Spatial Change Features] from a raw image followed by its normalization. The use of spatial change features is biologically inspired from the fact that in primates spatial change forms an essential visual cue for recognition [94]. The normalized spatial change features are then used in the LBDS classifier. In practice, face recognition algorithms are used to determine the class of a test face image from a set of gallery face images. This requires the comparison of the normalized spatial change features of the test image with normalized spatial change features of every image in the gallery.

Comparison is done by using a similarity measure. Thresholding this similarity measure enables local binary decisions that results in a binary decisions vector. As discussed in Chapter 2 and Section 3.2, the use of local binary decisions is biologically inspired from the firing of neurons that enable the process of decision making in primate brains. The binary decision vectors formed as a result of comparisons between a test image and a set of gallery images are used for similarity score calculation. This is followed by determination of its class by ranking the calculated similarity scores. Fig. 3.1 shows a brief overview of the
3. Local Binary Decisions Based Face Recognition

main components in the LBDS baseline algorithm.

### 3.4.1 Feature Extraction

Each pixel intensity in a raw image is represented by \( I(i, j) \), where \( (i, j) \) represents the location of a pixel in the image \( I \). The values of \( i \) and \( j \) are limited by the maximum number of rows \( M \) and the maximum number of columns \( N \), respectively. The dimensionality \( D \) of a raw image is equal to the scalar product of \( M \) and \( N \). Using such a raw image, spatial change features can be extracted by applying a standard local spatial filter. The feature extraction in the LBDS algorithm involves the following two steps: (a) spatial change feature extraction, and (b) normalization of the extracted spatial change features.

#### 3.4.1.1 Spatial Change Features

The texture of an image is represented by the surface intensities in the image. The relative change of spatial intensity of a pixel in a raw image with the corresponding pixels in its neighbourhood can be used to form features useful for recognition. In the baseline algorithm we can detect such features by calculating the local standard deviation on the image pixels encompassed by a window \( w \) of pixels of size \( m \times n \) pixels. This type of spatial operation is known as a kernel based local spatial filtering. The local standard deviation filter is given by the following equation:

\[
\sigma(i, j) = \sqrt{\frac{1}{mn} \sum_{s=-a}^{a} \sum_{t=-b}^{b} [I(i + s, j + t) - \overline{I(i, j)}]^2} \tag{3.1}
\]

where \( a = (m - 1)/2 \) and \( b = (n - 1)/2 \). The local mean \( \overline{I(i, j)} \) used in (3.1) is calculated by the following equation:

\[
\overline{I(i, j)} = \frac{1}{mn} \sum_{s=-a}^{a} \sum_{t=-b}^{b} I(i + s, j + t) \tag{3.2}
\]
In Fig. 3.2, the images labeled (a), (b), and (c) show the raw images, whereas the images labeled (d), (e), and (f) show the corresponding spatial change features [using Eq. (3.1)] respectively.

### 3.4.1.2 Normalized Spatial Change Features

The normalized spatial change features $\hat{x}$ are calculated using the following equation:

$$ x(i, j) = \frac{\sigma(i, j)}{\bar{\sigma}} $$

(3.3)

where the spatial change features $\sigma$ are normalized using the global mean $\bar{\sigma}$. The global mean is calculated with the following equation:

$$ \bar{\sigma} = \frac{1}{NM} \sum_{i=1}^{N} \sum_{j=1}^{M} \sigma(i, j) $$

(3.4)

In Fig. 3.2, the images labeled (g), (h), and (i) show the normalized spatial change features which is obtained by applying global-mean normalization on spatial change features images labeled (d), (e), and (f), respectively.

### 3.4.2 Local Binary Decisions on Similarity Classifier

Classifiers are required for comparing and classifying a test image $I_t$ with images in the gallery $I_g^{(k)}$, where $k$ represent the index of an image in the gallery of size $K$. The local binary decisions on similarity classifier has three essential steps: (a) normalized similarity calculation, (b) local binary decisions, and (c) similarity score calculation and ranking.

#### 3.4.2.1 Normalized Similarity Measure

Using Eqs. (3.1) and (3.3), normalized spatial change features vectors $x_t$ and $x_g^{(k)}$ are formed from $I_t$ and $I_g^{(k)}$ respectively. These feature vectors are then used for the comparison of a test image with the set of gallery images. Various
3. Local Binary Decisions Based Face Recognition

different distance measures and algorithms can be used to compute the similarity between two comparing features [96–100]. The absolute difference between pixels is a well known distance measure used for the comparison of features and can be used to find the similarity. Further, element wise normalization of this similarity measure is done by taking the minimum of each feature within $x_t$ and $x_g$ under comparison. This feature by feature comparison results in a normalized similarity measure $\delta^{(k)}$, which is given by:

$$\delta^{(k)}(i,j) = \frac{|x_g^{(k)}(i,j) - x_t(i,j)|}{\min(x_g^{(k)}(i,j), x_t(i,j))}$$  \hspace{1cm} (3.5)

The images labeled (j) and (k) in Fig. 3.2 are obtained by applying Eq. (3.5) to compare the normalized spatial change features labeled (h) with the normalized spatial change features labeled (g) and (i), respectively. It can be observed that (h) is the normalized spatial change features of the test image labeled (b), whereas (g) and (i) represent the normalized spatial change features formed from the gallery images (a) and (c), respectively. The darker areas in the images labeled (j) and (k) indicate a similarity and whiter areas indicate a dissimilarity between the comparing features.

3.4.2.2 Local Binary Decisions

The use of local binary decisions is inspired by the firing of neurons, where each pulse is produced based on a threshold and represents a decision on an analog information. Further, the use of digital one-off decisions at local level (here pixel based features) enables robust comparison under noisy environment. We achieve such decisions by thresholding the normalized similarity measure $\delta^{(k)}$ and form a binary decision vector $B_0^{(k)}$ from $\delta^{(k)}$ using a global threshold $\theta$ by applying the following transformation:

$$B_0^{(k)}(i,j) = \begin{cases} 
1 & \delta(i,j) < \theta \\
0 & \delta(i,j) \geq \theta 
\end{cases}  \hspace{1cm} (3.6)$$
3. Local Binary Decisions Based Face Recognition

The binary decisions when applied on the images labeled (j) and (k) in Fig. 3.2 using Eq. (3.6) result in binary decision vectors shown by the images labeled (l) and (m), respectively. Here the white color (or binary value of ’1’ in $B^0_k$) represents a match of two comparing features and hence greater the white area higher will be the similarity between the images under comparison.

### 3.4.2.3 Similarity Score and Ranking

To quantify the similarity between any two comparing images, the local decisions are accumulated by summing all elements of decision vector $B^0_k$. This results in a similarity score $S_g^{(k)}$, which is mathematically expressed as:

$$S_g^{(k)} = \sum_{i=1}^N \sum_{j=1}^M B^0_k(i,j)$$  \hspace{1cm} (3.7)

where $k$ corresponds to the index of a gallery image. A comparison of a single test image with $K$ gallery images will result in $K$ similarity scores $\{S_g^{(1)}, \ldots, S_g^{(k)}, \ldots S_g^{(K)}\}$. Ranking is performed by taking the highest value of these similarity scores. This determines the best match of the test image with the gallery images. This process is summarized as:

$$k^* = \arg \max_k S_g^{(k)}$$  \hspace{1cm} (3.8)

### 3.4.2.4 Localization Error Compensation

Localization errors can effect the performance of face recognition system. However, in most practical situations these errors are difficult to be quantified and are random in nature. To deal with this problem various methods to compensate for localization errors can be tried. In LBDS algorithm, one of the simple ways of compensating localization error is by the perturbation of test image $I_t$ to generate a set of $T$ test images $I_t^{t1}$, where $t1 = 1, 2, \ldots, T$. The comparison of a set of test images $I_t^{t1}$ with a gallery image $I_g^{(k)}$ results in $T$ similarity scores $S_g^{t1}$. The
maximum among these similarity scores represents the final similarity score $S_g^{(k)}$ for the comparison between the test image $I_t$ and and the gallery image $I_g^{(k)}$.

$$S_g^{(k)} = \max \left[ S_g^1, S_g^2, \ldots, S_g^{T-1}, S_g^T \right]$$  \hspace{1cm} (3.9)

A second method to compensate for the localization error is by the perturbation of gallery image $I_g^{(k)}$ to generate a set of $G$ test images $I_g^{(k)g1}$, where $g1 = 1, 2, \ldots, G$. The comparison of a test images $I_t^{(k)}$ with a set of generated gallery images $I_g^{(k)g1}$ of a person results in $G$ similarity scores $S_g^{g1}$. The maximum among these similarity scores represents the final similarity score $S_g^{(k)}$ for the comparison between the test image $I_t$ and and the gallery image $I_g^{(k)g1}$.

$$S_g^{(k)} = \max \left[ S_g^1, S_g^2, \ldots, S_g^{g1}, \ldots, S_g^{G-1}, S_g^G \right]$$  \hspace{1cm} (3.10)

### 3.5 Experimental Details

The LBDS algorithm is applied to AR [1], ORL [101], YALE [78], CALTECH [102], and FERET [103] standard face image databases. At any specific time, illumination, occlusions, face expressions, and time gap between the gallery and test images form variabilities that make the face recognition difficult. A difficult and practically important face-recognition task is created by limiting the gallery to a single image per person. Unless otherwise specified, the results presented in this letter are obtained by this kind of open-set testing.

For each image in the AR, YALE, and CALTECH databases, the eye coordinates of the face images are registered manually. For FERET database, the eye coordinates provided in the FERET distribution DVD is used for face alignment. The face alignment is done by rotating, shifting, and scaling the faces so that for all the faces the distance between the eyes remains constant and in fixed spatial coordinates. All the images were aligned and cropped to image size of
3. Local Binary Decisions Based Face Recognition

160×120.\(^2\) However, as ORL images are approximately localized images, manual alignment are not done on it and are resized to 40 × 32 pixels.

Since the eye coordinates of the faces in AR, Yale, and Caltech databases are detected manually they show shift errors after processing. The eye coordinates of the faces in the gray FERET database are provided within the FERET distribution DVD, and when used, show rotation and scaling errors. Perturbation to eye coordinates are done to compensate for these localization errors. These modifications are in the range of 1 to 6 pixels.

Unless otherwise specified, the following global settings are used for the set of LBDS parameters. To calculate spatial intensity change, the local standard deviation filter [see Eq. (3.1)] is used with optimal window size of 7 × 5 and 3 × 3 pixels when image size is 160 × 120 and 40 × 30 pixels respectively. The min-max similarity ratio shown in Table 3.1 is used. Finally, the value of the global threshold \( \theta \) is set to 0.7 which is selected empirically. The number of perturbation used for compensating localization errors in every case is set to a value of 15.

3.6 Results and Discussion

The overall recognition accuracy for the 2500 test images and the gallery size of 100 in the AR database is 91%. This very high accuracy level is possible due to the consistent performance over the large number of variable conditions that are individually listed in Table 3.2. Similar accuracy levels are obtained for YALE, ORL and CALTECH databases as shown in Table 3.3. As expected, increased variations correspond to decreased recognition accuracies in all databases. The demonstrated robustness of the algorithm is consistent with the fact that the baseline algorithm does not require any prior knowledge of the specific condition.

\(^2\)This is done using the Unix script provided for face normalization in the CSU Face Identification Evaluation System, Version 5.0 [104].
Table 3.2: Recognition performance of the LBDS algorithm (Single training sample per person problem)

<table>
<thead>
<tr>
<th>Test conditions</th>
<th>Recognition accuracy on AR database (%)</th>
<th>Localization error compensation</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Yes(^a)</td>
<td>No</td>
</tr>
<tr>
<td><strong>Session 1 images</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Expression</td>
<td>99</td>
<td>98</td>
</tr>
<tr>
<td>Illumination</td>
<td>97</td>
<td>94</td>
</tr>
<tr>
<td>Eye occlusion</td>
<td>100</td>
<td>100</td>
</tr>
<tr>
<td>Eye occlusion, Illumination</td>
<td>95</td>
<td>80</td>
</tr>
<tr>
<td>Mouth occlusion</td>
<td>97</td>
<td>93</td>
</tr>
<tr>
<td>Mouth occlusion, Illumination</td>
<td>93</td>
<td>86</td>
</tr>
<tr>
<td><strong>Session 2 images</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Neutral</td>
<td>99</td>
<td>96</td>
</tr>
<tr>
<td>Expression</td>
<td>86</td>
<td>80</td>
</tr>
<tr>
<td>Illumination</td>
<td>85</td>
<td>80</td>
</tr>
<tr>
<td>Eye occlusion</td>
<td>90</td>
<td>83</td>
</tr>
<tr>
<td>Eye occlusion, Illumination</td>
<td>77</td>
<td>62</td>
</tr>
<tr>
<td>Mouth occlusion</td>
<td>89</td>
<td>74</td>
</tr>
<tr>
<td>Mouth occlusion, Illumination</td>
<td>78</td>
<td>60</td>
</tr>
<tr>
<td><strong>Overall accuracy</strong></td>
<td>91</td>
<td>84</td>
</tr>
</tbody>
</table>

\(^a\) LBDS algorithm depicted here uses test image perturbations of ±5 pixels.

\(^b\) Results not available from the literature.
Table 3.3: Summary of the results on different databases

<table>
<thead>
<tr>
<th>Condition index a</th>
<th>Database b</th>
<th>Top rank recognition accuracy (%)</th>
<th>Localization error compensation</th>
</tr>
</thead>
<tbody>
<tr>
<td>(a)</td>
<td>CALTECH</td>
<td>89</td>
<td>No</td>
</tr>
<tr>
<td>(a)</td>
<td>YALE</td>
<td>93</td>
<td>No</td>
</tr>
<tr>
<td>(b)</td>
<td>ORL</td>
<td>72</td>
<td>No</td>
</tr>
<tr>
<td>(c)</td>
<td>FERET:Fb</td>
<td>85</td>
<td>Yes</td>
</tr>
<tr>
<td>(d)</td>
<td>FERET:Fc</td>
<td>71</td>
<td>Yes</td>
</tr>
<tr>
<td>(e)</td>
<td>FERET:Dup I</td>
<td>50</td>
<td>Yes</td>
</tr>
<tr>
<td>(f)</td>
<td>FERET:Dup II</td>
<td>40</td>
<td>Yes</td>
</tr>
</tbody>
</table>

a (a) Expression and illumination with a small gallery; (b) Small pose variation on small gallery (c) Expression on large gallery (Fb); (d) Illumination on large gallery (Fc); (e) Large gallery with mean time gap of 251 days (Dup I); (f) Large gallery with mean time gap of 627 days (Dup II).

b Single training image per person is used to form the gallery set. The sizes of the gallery sets are 28 in Caltech, 15 in YALE, 40 in ORL and 1196 in FERET databases; the sizes of the test sets are 150 in the YALE database, 406 in the CALTECH database, 360 in the ORL database, 1194 in set Fb, 194 in set Fc, 722 in Dup I, and 234 in Dup II of the FERET database.

that causes the dominant variations. To substantiate the claim of robustness, it is important to report the performance for a large gallery set. In practice, an increased gallery size decreases the overall recognition accuracy of any face recognition system. The results of testing with the FERET database, also shown in Table 3.3, demonstrate that the robustness is maintained under this condition.

Using the AR database, the effects of block size used to make the local binary decisions is analyzed and the results are shown in Fig. 3.3. The maximum recognition accuracy is achieved when the local binary decisions are made at the level of individual pixels (block size of one pixel) with a steep drop in the recognition accuracy as the block size is increased. This directly implies that larger image resolutions could further improve the recognition accuracy.

The impact of different implementations of the similarity measure is also analyzed. Using the implementations listed in Table 3.1, the change observed in the recognition accuracy is within 1%. Furthermore, the global threshold $\theta$ for making the local decisions is not a sensitive parameter. It is found that the recognition accuracy remains within 1% across various databases for a range of threshold values from 0.6 to 0.8. This confirms the general applicability of
Figure 3.3: The dependence of the overall recognition accuracy on the block size used to make the local binary decisions. The resolution of the images is 160×120 pixels. The window size of the standard-deviation filter is 7×5 pixels and the size of the normalization window is 80×60 pixels.

localised decisions on similarity as a concept.

The impact of the spatial change as features in the baseline algorithm are studied by using raw images as the feature vectors instead of spatial change feature vectors. The recognition accuracy for the AR database dropped from 91% to 63%. Furthermore, investigation on different filters for calculating the spatial intensity changes shows that the variation of the recognition accuracy with the standard local spatial filters: standard deviation, range and gradient, is within 1%. Based on this and the clear performance difference between the use of raw images and the spatial intensity changes as the feature vectors, it is concluded that the spatial intensity change is the visual cue for face recognition.
3. Local Binary Decisions Based Face Recognition

3.7 Conclusions

In this chapter, the local binary decisions is identified an important concept that is required for recognition of faces under difficult conditions. In addition, spatial intensity changes is identified as the visual cue for face recognition. A baseline algorithm, formed by implementing the local binary decisions based classifier and the spatial intensity changes based feature extractor, shows a robust performance under difficult testing conditions. To increase the recognition performance, a baseline system is formed by including perturbation scheme for localization error compensation. Using this baseline system the effect of localization errors is analysed. Further, the analysis shows that the application of the principles of local binary decisions and modularity results in a highly accurate face recognition system. The presented algorithm does not use any known configurational information from the face images, which makes it applicable to any visual pattern classification and recognition problem. Furthermore, classifiers based on the local binary decisions on similarity can be used in other pattern recognition applications.
Chapter 4

Analysis of LBDS Algorithm

4.1 Introduction

This chapter provides a detailed analysis and relevant comparison of Local Binary Decisions on Similarity (LBDS) algorithm which uses spatial intensity change based features and local binary decisions based classifier as its conceptual components. Further, the performance of this algorithm is evaluated when training set is limited to single training sample per person.

Section 3.4 described a baseline algorithm that forms the starting point for analysis of the LBDS algorithm. The analysis provided in Section 4.2 provides an insight into the individual components in the algorithm and its alternative implementations. Section 4.2 comprises of various experimental analysis followed by an implementation of the LBDS final algorithm. The final LBDS algorithm can be considered as a modular system as it contains different conceptual components such as: spatial change detection and local binary decisions. The optimal parameters required to properly implement the components in proposed method is done in Section 4.2. This finalized algorithm is then compared and discussed with other selected face recognition algorithms in Section 4.3. The final Section 4.4 summarizes the results obtained in this chapter.
4. Analysis of LBDS Algorithm

Analysis of the LBDS baseline algorithm enables to (1) understand the working of the algorithm and its components, (2) optimize the parameters, and (3) set all the specific details of the final LBDS algorithm that will enable to benchmark it with other algorithms. A detailed analysis is provided on the effects of (a) spatial change features in Section 4.2.2.1, (b) normalization of features and similarity measures in sections 4.2.2.2, 4.2.2.3, and 4.2.2.4, (c) local binary decisions and threshold in Section 4.2.2.5, (d) resolution of features and decision block size in Section 4.2.2.6, (e) color in Section 4.2.2.7, and (f) localization errors in Section 4.2.2.8. The images from AR face database is used for the analysis and experiments presented in this section.

4.2 Experimental Analysis of the Algorithm

Analysis of the LBDS baseline algorithm enables to (1) understand the working of the algorithm and its components, (2) optimize the parameters, and (3) set all the specific details of the final LBDS algorithm that will enable to benchmark it with other algorithms. A detailed analysis is provided on the effects of (a) spatial change features in Section 4.2.2.1, (b) normalization of features and similarity measures in sections 4.2.2.2, 4.2.2.3, and 4.2.2.4, (c) local binary decisions and threshold in Section 4.2.2.5, (d) resolution of features and decision block size in Section 4.2.2.6, (e) color in Section 4.2.2.7, and (f) localization errors in Section 4.2.2.8. The images from AR face database is used for the analysis and experiments presented in this section.

4.2.1 Database

The AR face database [1,2] contains over 4,000 color face images of 126 people (70 men and 56 women), which include frontal pose of faces with different facial expressions, illumination, and occlusions. 100 people (50 men and 50 women) are randomly selected for the study whose photos were taken in two sessions (separated by two weeks). Each session contains 13 color images that correspond
to the following conditions: 1 neutral expression, 3 expressions, 3 illuminations, 2 occlusions without illumination and 4 occlusions with illumination. Manual cropping is done on the face portion of the image and followed by normalization to 160 × 120 pixels. Examples of aligned session-one and session-two images from the AR database are shown in Fig. 4.1(a-m) and Fig. 4.1(n-z), respectively. For the experiments presented in this chapter using the AR database, 100 neutral face images of unique individuals from the first session [shown as the image with label (a) in Fig. 4.1] are used for forming the gallery set, whereas the remaining 2500 face images from both sessions [shown as the images with labels from (b) to (z) in Fig. 4.1] form the test set.

4.2.2 Analysis and Discussion

Unless specified otherwise, all the experiments presented in this section are conducted using the AR face database with the following numerical values: 0.25 for \( \theta \), 160 × 120 pixels for the image size, and 7 × 5 pixels for the kernel window size of the standard deviation filter.

4.2.2.1 Effect of Spatial Intensity Change Used as Features

It is a well known fact that the edges of the objects in an image define its shape. This knowledge has been widely applied in image processing studies for segmentation, shape detection, and texture analysis [105–119]. However, this knowledge is not widely used in face recognition studies. An analysis using spatial change features and raw features suggest that inter pixel spatial change within an image is the essential photometric or geometric visual cue that contributes to the recognition of the objects in it. This can be observed from the results presented in Table 4.1.

The analysis of the performance using various features with and without mean normalization is shown in Table 4.1. The importance of spatial change as
features for face recognition is analysed by comparing its performance with raw features and edge features. For this comparison the standard nearest neighbour (NN) classifier [120–123] and the LBDS classifier are used.

A raw face image in itself contains all the identity information required for face recognition. However, occurrence of external occlusions, expressions, and illumination in face images can result in loss of such identity information. Further, raw image intensities are highly sensitive to variations in illumination, which make recognition on raw images a difficult task. The comparison shown in Table 4.1 between spatial change features and raw image features clearly shows that spatial change features outperforms the raw features significantly. This superior performance of spatial change features over raw features can be attributed to the facts that spatial change features (1) shows lower local variability in the face images under various conditions such as expression, illumination, and occlusion, and (2) preserves the identity information of a face.

Most edge detection techniques are inaccurate approximations of image gradients. Spatial change detection techniques are different from standard edge detection techniques. The majority of the edge detection techniques result in the removal of medium to small texture variations as distinct from spatial change detection techniques that preserve most of the texture details. Such variations however contain useful information for identification and show increased recognition performance. These observations are shown in Table 4.1. They further confirm the superiority of using spatial change features in face recognition and show the relative difference of using spatial change features as opposed to the edge features.

Figure 4.2 is a graphical illustration of the overall impact of using spatial change features. The plot shows a normalized histogram of similarity scores $S_g^{(k)}$ [see Eq. (3.7)] resulting from inter-class and intra-class comparisons. The 100 gallery images described in Section 4.2.1 form the 100 classes and are compared
### Table 4.1: Effect of global mean normalization and feature type

<table>
<thead>
<tr>
<th>Index</th>
<th>Feature Type</th>
<th>Recognition accuracy (%)&lt;sup&gt;a&lt;/sup&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>NN Classifier</td>
</tr>
<tr>
<td><strong>With global mean normalization&lt;sup&gt;a&lt;/sup&gt;</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Raw features</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>r1</td>
<td>Raw</td>
<td>46.0</td>
</tr>
<tr>
<td><strong>Spatial change features</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>s1</td>
<td>Local Standard Deviation</td>
<td>67.6</td>
</tr>
<tr>
<td>s2</td>
<td>Local Range</td>
<td>68.6</td>
</tr>
<tr>
<td><strong>Edge</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>e1</td>
<td>Sobel edges</td>
<td>69.0</td>
</tr>
<tr>
<td>e2</td>
<td>Prewitt edges</td>
<td>69.2</td>
</tr>
<tr>
<td><strong>Without global mean normalization</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Raw features</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>r2</td>
<td>Raw</td>
<td>38.5</td>
</tr>
<tr>
<td><strong>Spatial change features</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>s1</td>
<td>Local Standard Deviation</td>
<td>59.3</td>
</tr>
<tr>
<td>s2</td>
<td>Local Range</td>
<td>63.0</td>
</tr>
<tr>
<td><strong>Edge</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>e1</td>
<td>Sobel edges</td>
<td>50.4</td>
</tr>
<tr>
<td>e2</td>
<td>Prewitt edges</td>
<td>49.4</td>
</tr>
</tbody>
</table>

<sup>a</sup> Global mean normalization is achieved using Eq. (3.3) and Eq. (3.4). While for raw features normalization is done by replacing \(\sigma(i, j)\) with \(I(i, j)\) in Eq. (3.3) and Eq. (3.4).

<sup>b</sup> LBDS baseline algorithm with global mean normalization.
Figure 4.2: Graphical illustrations showing the overall influence of using spatial change features. The graphs show a normalized frequency distribution of similarity scores $S_g$ when using (a) spatial intensity change features (b) raw image features.
against 2500 test images in the AR database. The inter-class plots are obtained by comparing each of these test images with the gallery images belonging to a different class, whereas intra-class plots are obtained by the comparison of each test image against a gallery image belonging to its own class. Further, a comparison is done between spatial change features (Fig. 4.2a) and raw image features (Fig. 4.2b). The overlapping region of the two distributions indicate the maximum overall probability of error when using the LBDS classifier. This region also shows the maximum overall false acceptance and false rejection that can occur in the system. A smaller area of overlap implies better recognition performance. Clearly, it can be seen that the use of feature vectors in Fig. 4.2a as opposed to the raw-image features in Fig. 4.2b results in a smaller region of overlap and hence better recognition performance.

An analysis is done to study the effect of using a spatial change filter window \( w \) of various sizes \([w \text{ is described in Section (3.4.1.1)})\]. It can be observed from Fig. 4.2 that with an increase in resolution of the spatial change features (or the raw image) the recognition performance shows increased stability against variation in spatial change filter window size. Further, it can be also seen that higher resolution images show better recognition accuracies.

### 4.2.2.2 Normalization

The baseline algorithm contains two different types of normalization. They are: (1) global mean normalization of the feature vectors as described in Section 3.4.1.2 and (2) similarity measure normalization employed in the classifier, which is described in Section 3.4.2.1. The relative importance of using these normalization methods is presented in Table 4.2. It is observed that normalization of the distance measures results in higher recognition accuracies. It can also be observed that global mean normalization shows improved recognition accuracy only when similarity measure normalization is used, which also shows that global
4. Analysis of LBDS Algorithm

Figure 4.3: A graphical illustration showing the recognition performance of the LBDS algorithm under the variation of spatial change features filter window size at various image resolutions.

Table 4.2: Effect of Global Mean Normalization of Features and Similarity Measure Normalization

<table>
<thead>
<tr>
<th>Condition</th>
<th>Features</th>
<th>Similarity measure</th>
<th>Recognition accuracy (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>NN Classifier</td>
<td>LBDS Classifier</td>
</tr>
<tr>
<td>(a)</td>
<td>Yes</td>
<td>Yes</td>
<td>67.6</td>
</tr>
<tr>
<td>(b)</td>
<td>Yes</td>
<td>No</td>
<td>67.6</td>
</tr>
<tr>
<td>(c)</td>
<td>No</td>
<td>Yes</td>
<td>59.3</td>
</tr>
<tr>
<td>(d)</td>
<td>No</td>
<td>No</td>
<td>59.3</td>
</tr>
</tbody>
</table>

a Feature extraction filter window used in Eq. (3.2) has a size of 7 × 5 pixels for a raw image I with a size of 160 × 120 pixels. Normalized similarity measure described using Eq. (3.5) is used for these simulations.

b The results are shown for the best accuracies by optimizing the threshold \( \theta \). The optimized values of the threshold for the condition indexes (a), (b), (c) and (d) are 0.5, 0.25, 0.35 and 0.85 respectively.
4. Analysis of LBDS Algorithm

Mean normalization in isolation does not improve the recognition performance. In the following sections the effect of these two normalization is further studied and alternative methods are attempted. This is done to provide a better technical insight into the normalization methods. This also helps in understanding the unique features that contribute to the overall recognition performance.

4.2.2.3 Effect of Mean Normalization and Study of Alternative Normalization

From the experimental results obtained in Table 4.2, it is found that the normalization of spatial change features by a global mean is not effective in improving the recognition performance. Clearly, the feature normalization performed by Eq. (3.3) does not improve the performance considerably, which leads us to investigate alternative local mean normalization techniques. Equation (3.4) is now replaced by the following equation to calculate the local mean of spatial change features:

$$\sigma(i,j) = \frac{1}{kl} \sum_{s=-a1}^{a1} \sum_{t=-b1}^{b1} \sigma(i+s,j+t)$$  (4.1)

where the moving window of pixels is of size $k \times l$ pixels, $a1 = (k - 1)/2$ and $b1 = (l - 1)/2$. Local mean normalization is applied on spatial change features by using Eq. (4.1) followed by Eq. (3.3).

An investigation on the performance of using local mean normalization with local mean windows of different sizes is done. Figure 4.4 shows the effect of variation in local mean window on the recognition performance when using spatial change features and raw features. Further, the same graph shows a comparison of its performance with global mean normalization. It is observed that recognition performance increases when features are normalized using the local mean normalization described by Eq. (4.1) and Eq. (3.3). The improvement in recognition accuracy while using local mean normalization as distinct from global mean normalization is relatively large in the case of the raw features while hav-
Figure 4.4: Graphical illustration showing improved performance of local mean normalization compared to global mean normalization. The graph show the following conditions: (a) local mean normalization applied to raw features, (b) local mean normalization applied to spatial change features, (c) global mean normalization applied to raw features, and (d) global mean normalization applied to spatial change features. The image size is 160 × 120 pixels; \( w \) is of size 7 × 5 pixels; the local mean filter window size is varied from 10 × 7 pixels to 160 × 120 pixels; for each local mean filter window size the best recognition accuracy is selected by optimizing the threshold. Normalized similarity measure given by Eq. (3.5) is used for these simulations.
4. Analysis of LBDS Algorithm

Figure 4.5: Graphical illustration showing the effect of local mean normalization and similarity measure normalization on the performance of the LBDS algorithm. The graph show the following conditions: (a) local mean normalization applied to spatial change features and with normalization similarity measure for comparison, (b) local mean normalization applied to spatial change features and with similarity measure without normalization for comparison, (c) spatial change features without normalization and with normalized similarity measure comparison, and (d) spatial change features without normalization and with similarity measure without normalization for comparison. Normalization of features is performed using global mean normalization of spatial change features using Eq. (3.4) and Eq. (3.3). This feature normalization is tried in combination with normalized similarity measure and the performances are compared.
Figure 4.6: Graphical illustration showing the effect of global mean normalization and similarity measure normalization on the performance of the LBDS algorithm. The graph show the following conditions: (a) global mean normalization applied to spatial change features and with normalization similarity measure for comparison, (b) global mean normalization applied to spatial change features and with similarity measure without normalization for comparison, (c) spatial change features without normalization and with normalized similarity measure comparison, and (d) spatial change features without normalization and with similarity measure without normalization for comparison. Normalization of features is performed using global mean normalization of spatial change features using Eq. (3.4) and Eq. (3.3). This feature normalization is tried in combination with normalized similarity measure and the performances are compared.
ing very little impact on spatial change features. Further, in comparison with the raw features, the spatial change features is stable for a broader range of local mean normalization filter window size. The algorithm using spatial change features provides robust performance within the local mean normalization filter window range of $80 \times 60$ pixels to $40 \times 30$ pixels as shown in Fig. 3.

Table 4.3 shows the effect of using local mean normalization on spatial change features. Clearly, in comparison with Table 4.2, the local mean normalization on spatial change features shows an increase in recognition performance when using the LBDS classifier. However, the recognition performance shows no improvement when using an NN classifier. Further, Fig. 4.5 shows that local mean normalization improves the overall recognition performance and provides a wider stable range of threshold than when using global mean normalization [see Fig. 4.6 and Fig. 4.5]. It can be observed that in comparison with global mean normalization on similarity measure, the local mean normalization on similarity measure shows increased stability in recognition accuracy with respect to a varying threshold. All these effects make local mean normalization the preferred choice for use in a feature normalization process.

### 4.2.2.4 Effect of Similarity Measure Normalization and Study of Alternative Normalization

Normalization of the similarity measures also helps in increasing the recognition accuracy of the LBDS algorithm and enables a stable threshold. This is evident from (1) Table 4.2 and Table 4.3, showing the superiority of similarity measure normalization over mean normalization techniques and (2) Fig. 4.5 and Fig. 4.6 showing the relative importance of similarity measure normalization in stabilizing the threshold range and increasing the recognition performance. Further, the improvement of recognition performance provided by normalizing the similarity measure can be observed from Table 4.4. It can be observed that all of the
4. Analysis of LBDS Algorithm

Table 4.3: Effect of Local Mean Normalization and Distance Normalization

<table>
<thead>
<tr>
<th>Condition</th>
<th>Features</th>
<th>Similarity measure</th>
<th>Recognition accuracy (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>(a)</td>
<td>Yes</td>
<td>Yes</td>
<td>62.0 86.2</td>
</tr>
<tr>
<td>(b)</td>
<td>Yes</td>
<td>No</td>
<td>62.0 81.9</td>
</tr>
<tr>
<td>(c)</td>
<td>No</td>
<td>Yes</td>
<td>59.3 84.7</td>
</tr>
<tr>
<td>(d)</td>
<td>No</td>
<td>No</td>
<td>59.3 78.4</td>
</tr>
</tbody>
</table>

a Feature extraction filter window used in Eq. (3.2) has a size of $7 \times 5$ pixels for a raw image $I$ with a size of $160 \times 120$ pixels. The size of local mean normalization window $w_1$ used in Eq. (4.1) is set to $80 \times 60$ pixels. Normalized similarity measure described using Eq. (3.5) is used for these simulations.

b The results are shown for the best accuracies by optimizing the threshold $\theta$. The optimized values of the threshold for the normalization conditions (a),(b),(c) and (d) are 0.5, 0.25, 0.35 and 0.85 respectively.

Table 4.4: Direct and Normalized Similarity Measures

<table>
<thead>
<tr>
<th>Index</th>
<th>Similarity measurea</th>
<th>Recognition accuracy (%)b</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Normalized</td>
<td></td>
</tr>
<tr>
<td>n1</td>
<td>$\frac{\min(x_g, x_t)}{\max(x_g, x_t)}$</td>
<td>85.9</td>
</tr>
<tr>
<td>n2</td>
<td>$\frac{</td>
<td>x_g - x_t</td>
</tr>
<tr>
<td>n3</td>
<td>$\frac{</td>
<td>x_g - x_t</td>
</tr>
<tr>
<td>n4</td>
<td>$\frac{</td>
<td>x_g - x_t</td>
</tr>
<tr>
<td>n5</td>
<td>$e^{-\frac{</td>
<td>x_g - x_t</td>
</tr>
<tr>
<td>n6</td>
<td>$e^{-\frac{</td>
<td>x_g - x_t</td>
</tr>
<tr>
<td>n7</td>
<td>$e^{-\frac{</td>
<td>x_g - x_t</td>
</tr>
<tr>
<td></td>
<td>Direct</td>
<td></td>
</tr>
<tr>
<td>d1</td>
<td>$</td>
<td>x_g - x_t</td>
</tr>
<tr>
<td>d2</td>
<td>$e^{-</td>
<td>x_g - x_t</td>
</tr>
</tbody>
</table>

a Feature extraction filter window used in Eq. (3.2) has a size of $7 \times 5$ pixels for a raw image $I$ with a size of $160 \times 120$ pixels. The size of local mean normalization window $w_1$ used in Eq. (4.1) is set to $80 \times 60$ pixels.

b $\theta$ is optimised for highest accuracies on each similarity measure under consideration.
4. Analysis of LBDS Algorithm

Figure 4.7: Graphical illustration showing a comparison of normalized similarity measure with a direct similarity measure. The image size is $160 \times 120$ pixels; the size of $w$ is $7 \times 5$ pixels; the size of local mean filter window $w_1$ is set to $80 \times 60$ pixels.

Normalized similarity measures outperform the corresponding direct similarity measures with respect to the recognition accuracy. Fig. 4.7 shows the influence of variable threshold on the normalized and direct similarity measures. Clearly, for every threshold the normalized similarity measures show better recognition performance than those without similarity measure normalization. These results suggest that normalization of similarity measures is an important factor that helps in improving the recognition performance of the LDBS algorithm.

4.2.2.5 Effect of Local Binary Decisions and Threshold

Binary decisions are made by transforming the normalized similarity measure to a binary decision vector by using a predefined global threshold. A threshold $\theta$ is used to set similar features to a value of one, whereas dissimilar features are set to a value of zero. The LBDS classifier applies the binary decisions to individual
pixels, which means that it can utilize the maximum available spatial change features in the image. To our knowledge, the only other method that attempted the binary decisions for face recognition is the ARENA face recognition algorithm [124], which used $L_0^*$ distance [where $L_0^*(\bar{a}) \equiv \sum |a_i|^p$] as its similarity measure, applied to the raw features. However, ARENA used low resolution images [16 × 16 pixels], which led the authors to a conclusion that the dimensionality reduction and noise reduction were the main contributing factors to the performance of that algorithm. When using low resolution images, the decisions are no longer made on individual pixels, instead binary decisions are effectively made on pixels that represent an average of a group of pixels. Therefore, ARENA does not use local binary decisions. Furthermore, ARENA does not use the spatial change features. As distinct from this, the LBDS classifier employs local binary decisions on spatial change features at the pixel level.

The importance of local binary decisions in the LBDS classifier is shown
4. Analysis of LBDS Algorithm

in Fig. 4.8. The comparison of recognition performance with thresholding and without thresholding shows a very large change from 86.2% to 13.8% respectively. This shows the relative importance of local binary decisions, confirming it as the essential component of the algorithm. The local binary decisions result in the removal of noisy information associated with the natural variability. For example, occlusions and facial expressions remove identity information from the face and can also add information that may seem to be relevant (false similarity) to a non-binary classifier such as the NN classifier. Without the binary decisions, the noisy information gets accumulated when forming a global similarity score (note that similarity scores are formed by adding the values of the elements in the similarity measure vector). Since the global similarity score has significant contribution of such noisy information (or false similarity), the result is a reduced recognition performance. As opposed to this, every feature is used for making local decisions in the case of the LBDS classifier. In this case, the global similarity score does not accumulate the effect of less similar features, resulting in a better recognition performance.

Figure 4.9 shows the performance of the LBDS algorithm with a change in threshold when using various normalized similarity measures. We can observe that the recognition accuracy is stable over a broad range of threshold values irrespective of the normalized similarity measures employed. The stability of the threshold and increased recognition performance can be attributed to the use of normalized similarity measures [see Fig. 4.7]. Further, the stability of the threshold enables the use of any of the possible similarity measures to form the LBDS classifier. A stable threshold in turn implies that the recognition performance of the algorithm is least sensitive to threshold variation. Further, this allows for the use of a single global threshold across different databases containing images of various types of natural variability.
Figure 4.9: Graphical illustration showing the stability of the threshold against various normalized similarity measures. The image size is 160 × 120 pixels, the size of the standard deviation filter is 7 × 5 pixels, and the value of the global threshold $\theta$ is varied from 0.1 to 0.9.
Figure 4.10: Graphical illustration showing the recognition performance of the LBDS algorithm with variation in resolution of the normalized similarity measure $\delta$ under comparison. Averaging is performed to reduce the resolution of $\delta$. 

*Average before* shows the case when raw images at various resolutions are used, whereas *average after* shows the case when spatial change features at various resolutions are formed from a 160 × 120 pixels raw image.
4.2.2.6 Effect of Resolution

The recognition performance with respect to variation in resolution can be studied by (1) varying the raw image resolution and (2) increasing the decision block size. In the first case, reducing the image resolution from a higher resolution will result in a smaller number of normalized spatial change features. The reduction of a higher resolution image to a lower resolution image can be achieved by averaging a block of pixels to form a single pixel. This averaging results in a loss of features and hence it is natural to expect that recognition performance will drop with lower resolution images and hence fewer features. We can observe from Fig. 4.10 that with lower resolution images the recognition performance drops considerably (this situation is labeled as average before).

In the second case, the resolution of spatial change features are kept to a maximum of 160 × 120 pixels, followed by the calculation of δ. The reduction in resolution is achieved by averaging on a block of elements in δ. Block by block reduction across the entire δ results in a lower resolution of δ. This situation is labeled as average after in Fig. 4.10. We can observe from Fig. 4.10 that in the case of average after, the reduction in resolution results in a slight reduction of the recognition performance, which however, again shows that a larger number of features helps to increase the recognition performance.

Figure 4.10 further shows the importance of having a larger number of features irrespective of the decision block size. A larger number of features and a smaller decision block size results in increased recognition performance. Further, as observed from Fig. 4.3, an increased resolution of features extends the stable range of spatial change filter window size.

4.2.2.7 Effect of Color

Color images are formed of three channels, namely, red, green, and blue. Table 4.5 shows that the use of color images also helps to improve the recognition
4. Analysis of LBDS Algorithm

Table 4.5: Effect of color on single training samples per person scheme

<table>
<thead>
<tr>
<th>Color combination</th>
<th>Recognition accuracy (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>AR (b)-(z)</td>
</tr>
<tr>
<td>c1 Gray</td>
<td>86.16</td>
</tr>
<tr>
<td>c2 Red</td>
<td>68.86</td>
</tr>
<tr>
<td>c3 Green</td>
<td>86.00</td>
</tr>
<tr>
<td>c4 Blue</td>
<td>87.64</td>
</tr>
<tr>
<td>c5 Red+Green</td>
<td>81.55</td>
</tr>
<tr>
<td>c6 Blue+Green</td>
<td>88.96</td>
</tr>
<tr>
<td>c7 Red+Blue</td>
<td>85.84</td>
</tr>
<tr>
<td>c8 max(c5,c6,c7)</td>
<td>89.60</td>
</tr>
</tbody>
</table>

a Similarity score calculated from (c1) gray images, (c2) red channel alone, (c3) green channel alone, (c4) blue channel alone, (c5) combination of scores from red and green channels, (c6) combination of scores from blue and green channels, (c7) combination of scores from red and blue channels, and (c8) the maximum of scores obtained as a result of operations c5 to c7.

Performance. Similarity scores for a comparison between a color test image and a color gallery image can be obtained by one-to-one comparison of red, green, and blue channels of one image to the other. To obtain an overall similarity score, an additive combination of the independent similarity scores observed across the red, green, and blue channels are taken. Table 4.5 lists some of the combinations that are used in our analysis. Table 4.5 further illustrates that the use of independent channels alone are not sufficient for robust performance. It can be also observed that utilizing the additive combination of similarity scores obtained from the channels of color images provides a higher recognition accuracy than when using gray images. This can be seen from the recognition performance of the LBDS algorithm when using the combination of the color channels (see c8 listed in Table 4.5). Although several other combinations can also be tried, analysis is limited to the extend to form a simple model for color, which is achieved through c8 listed in Table 4.5.
4.2.2.8 Effect of Localization

Automatic face detection and alignment is a difficult problem when natural variability in images is high. In any method that is based on pixel-by-pixel comparisons, it is essential that the features of the compared images are well aligned. Irrespective of the face detection method employed, natural variability can cause pixel-level misalignments. To compensate for the localization errors that occur after an automatic or manual alignment, we apply either test or gallery image shifts with respect to a set of registration points in the feature vectors. For example, the localization of face images can be achieved by detecting the location of eye coordinates. An error in localization means the eye coordinates are shifted. A scale error means that the eye coordinates are shifted towards each other or away from each other. A rotation error causes shifts of the two eye coordinates in opposite vertical directions. We perturbate the reference eye coordinates by applying such shifts and re-localize the face images using the shifted eye coordinates.

Using the above mentioned idea, two techniques that can be employed to reduce localization errors in the LBDS algorithm are (a) application of modifications such as shift, rotation, and scaling on the test image, followed by comparison with gallery, and (b) perturbation of the eye-coordinates of the gallery images to form several sets of synthetic gallery images. In both cases, each comparison of a test image with a gallery image, results in a similarity score $S^*_g$ that is calculated using Eq. (3.7) described in the baseline algorithm. The final similarity score $S_g$ for the test image with a compared gallery image is found by selecting the maximum $S^*_g$. Table 4.6 shows the recognition performance using both techniques using color and gray scale images. For these simulations the values of number of perturbations is set to 15, composed of 5 horizontal, 5 vertical and 5 diagonal perturbations. This performance difference is due to the fact that modification of test images is performed after cropping and results in
Table 4.6: Effect of Localization Error Compensation

<table>
<thead>
<tr>
<th>Color image</th>
<th>Recognition Accuracy (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Perturbation</td>
</tr>
<tr>
<td></td>
<td>No</td>
</tr>
<tr>
<td>Test image</td>
<td>89.6</td>
</tr>
<tr>
<td>Gallery image</td>
<td>86.2</td>
</tr>
</tbody>
</table>

loss of useful spatial information during comparison. This is different from the perturbation of the gallery images that preserves all the information from the original image.

4.2.3 LBDS algorithm

Based on the analysis presented in Section 4.2.2, LBDS algorithm that includes a method to compensate for localization errors and can use color images to improve the recognition performance is summarized in the following section.

4.2.3.1 Feature Vectors

The superiority of using spatial change features in comparison with raw features and edge features is shown through the analysis in Section 4.2.2.1. The spatial change features \( \sigma \) is determined from the raw image \( x \) using the Eq. (3.1). This is followed by calculating normalized spatial change features using Eq. (4.1) and Eq. (3.3). The only parameter that affects the formation of the spatial change features is the filter window size which is analysed in Fig. 4.3. For an image resolution of 160 \( \times \) 120 pixels, the spatial change features filter size that will be used is 7 \( \times \) 5 pixels. The need for local mean normalization over global mean normalization was analysed in Section 4.2.2.3, and the optimal local mean normalization window size for an image resolution of 160 \( \times \) 120 pixels is 80 \( \times \) 60 pixels. In general, the local mean normalization window size is kept at 0.25 times
that of the image resolution.

4.2.3.2 Determination of Similarity

The similarity between test image features $x_t$ and gallery image features $x_g^{(k)}$ is calculated using Eq. (3.5) followed by Eq. (3.6) and Eq. (3.7). Equation (3.5) is one of the normalized similarity measures analysed in Table 4.4. It can be noted that any of the normalized similarity measures shown in Table 4.4 could be used. However, Eq. (3.5) is selected as it provides a broader stable threshold than other equations listed in Table 4.4 and shown in Fig. 4.9. For the final algorithm the value of threshold $\theta$ is set to 0.25.

4.2.3.3 Use of Color Images

The feature vectors from red, green, and blue channels of the image are computed independently for the test and the gallery images. The similarity measure is used to calculate $\delta$ for each corresponding channel, in other words, the red, green, and blue channels of the test-image features are compared with the red, green, and blue channels of the gallery-image features, respectively. This is followed by the similarity score calculation using Eq. (3.6) for the comparisons between the red, green, and blue channels. The final similarity score for the comparison between the two images is calculated using c8 listed in Table 4.5.

4.2.3.4 Compensating Localization Errors

The study on localization errors performed in Section 4.2.2.8 reveals that such compensation also contributes towards increasing the recognition performance. Perturbations of either the test or the gallery images increase the performance, as shown in the Table 4.6. For the following sections, the method of perturbation of eye coordinates in the gallery images to compensate for shifting, scaling, and rotation errors is selected. Each generated image is treated independently to
4. Analysis of LBDS Algorithm

<table>
<thead>
<tr>
<th>Image size</th>
<th>w size</th>
<th>w1 size</th>
<th>θ</th>
</tr>
</thead>
<tbody>
<tr>
<td>160 × 120</td>
<td>7 × 5</td>
<td>80 × 60</td>
<td>0.25</td>
</tr>
<tr>
<td>170 × 150</td>
<td>7 × 5</td>
<td>85 × 75</td>
<td>0.25</td>
</tr>
</tbody>
</table>

calculate the similarity score $S_g^{(g1)}$ or $S_g^{(t1)}$ by using either gray images or channels of color images, where $t1$ and $g1$ is the number of applied perturbations per test image and gallery image respectively. The value of $t1$ or $g1$ is set to 15, formed by 5 horizontal shifts, 5 vertical shifts, and 5 diagonal shifts. The final similarity score $S_g$ is found by selecting the maximum of $S_g^{g1}$ or $S_g^{t1}$.

4.2.3.5 Parameter Settings

As can be seen from Table 4.7, the local mean normalization window size is set to $80 \times 60$ pixels (1/4th of image size) when using an image size of $160 \times 120$ pixels. The size of spatial change features filter is selected based on the analysis shown in Fig. 4.3. The value of global threshold $\theta$ is set to a constant value of 0.25. The total number of synthetic images generated, or the number of perturbations is set to 15.

4.3 Comparison with Other Methods

4.3.1 Databases

4.3.1.1 AR Database

The AR database described in Section 4.2.1 is modified to a lower resolution to provide a fair comparison with other methods described in the literature. The resolution of images in the AR database is reduced from size of $160 \times 120$ pixels to $66 \times 48$ pixels. The database formed with the lower resolution images is now
called as ARR database.

4.3.1.2 Gray FERET Database

Prior to processing, the faces are registered and aligned based on the markup data provided in the FERET distribution. In this study, only the head-on images are used; faces in profile or at other angles are discarded (see [103, 125, 126]). The terminology used for describing various partitions are adopted from [103, 125, 126]. For the study of large class problem, the gallery set is formed of 1,196 face images of unique subjects from Fa set [103, 125, 126]. The test set is formed of 1,195 images from probe Fb, 194 images from probe Fc, 722 images from probe Dup1 and 234 images from probe Dup2. In this work, the images were cropped to \(170 \times 150\) pixels.

4.3.2 Other Algorithms and Their Settings

The dimensionality of the face images is reduced based on the methods that are employed for face recognition. Most methods work only with smaller dimension. For PCA [10–12, 23, 26] based methods, the maximum possible dimensionality is equal to \(K - 1\), i.e maximum number of classes minus one. For the other methods, the dimensionality is decided by optimizing it for the best performance.

4.3.2.1 PCA

In our experiments with the AR database, gallery images of 100 persons (hence 100 classes) are used, which forces the maximum dimensionality to be fixed at 99.

4.3.2.2 PC\(^2\)A

This is an improved method over PCA and dimensionality is set same as that used in PCA. Further, the value of parameter \(\alpha\) described in [23] is set to 0.25.
Table 4.8: Comparison of Complexity

<table>
<thead>
<tr>
<th>Algorithm</th>
<th>Classification time</th>
<th>Storage space</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Gallery Perturbate</td>
<td>Test Perturbate</td>
</tr>
<tr>
<td>LBDS</td>
<td>$O(3shD)$(^a)</td>
<td>$O(3shD)$</td>
</tr>
<tr>
<td>SIS</td>
<td>$O(osEh^2v^2)$(^b)</td>
<td>$O(osEh^2v^2)$</td>
</tr>
<tr>
<td>LBDS</td>
<td>$O(3shD)$</td>
<td>$O(sD)$</td>
</tr>
<tr>
<td>SIS</td>
<td>$O(sh^2v^2D)$</td>
<td>$O(sD)$</td>
</tr>
</tbody>
</table>

\(^a\) In LBDS $h = v$, total number of synthetic image generated per person are $3h$ and the block size $E$ is set to 1. Further, $s$ in case of LBDS is much lower than SIS due the use of fewer number of spatial filters.

\(^b\) It can be noted that $cE > D$.

4.3.2.3 SVD PCA

This is yet another improved method over PCA and the dimensionality is set to 99. Further, the value of the parameter $\alpha$ described in [23] is set to 0.25 and that of $n$ described in [23] is set to 5/4.

4.3.2.4 Other Methods

The results of the following methods are directly used from the literature: SIS [24], $SIS_{nondiv}$ [24], LPS [28], SOM-face [25], EBGM [19] and LDA [14, 127].

These face recognition algorithms are the benchmark algorithms for the single training image per person problem. For all experiments on the PCA based methods described in sections 4.3.2.1, 4.3.2.2, and 4.3.2.3 NN classifier is used [128].

4.3.3 Computational Complexity

In this section, the classification time and storage complexity of the various algorithms are analysed. The terms used in the analysis are listed as follows:

$D$ Dimensionality of the image (in pixels)
s  Total number of training samples

h  Total number of horizontal perturbations

v  Total number of vertical perturbations

E  Decision block size (in pixels)

o  Number of overlapping decision partitioned regions [24] of size $E$

f  Reduced dimensionality of features, $f << D$

Table 4.8 shows the time complexity and storage requirements of LBDS in comparison with SIS and PCA. This table also shows the effect of perturbing the test or gallery images to form synthetic images (or perturbated features) for use in a classifier. In comparison with SIS, the LBDS algorithm has faster classification times and lower storage requirements. It can be observed that use of synthetic samples in the gallery increases the classification times and storage requirements of all the algorithms. In the specific case when only the test images are perturbated, the storage requirements are same as without perturbations. Although the LBDS algorithm performs better when the gallery is perturbated, it should be noted that perturbation of the test images yields similar results (see Table 4.6). The ability of LBDS to use test perturbation without a large drop in recognition accuracy is useful from a practical point of view, especially when working with a very large number of gallery face images, which limits the storage space. These results suggest the usefulness of the LBDS algorithm in a wide range of practical applications.

### 4.3.4 Experiments on AR

First, a comparison on the performances of the LBDS algorithm on the AR and ARR database is done. For this, 100 neutral face images of unique individuals
Table 4.9: Performance of the LBDS Algorithm with a Single Training Sample per Person

<table>
<thead>
<tr>
<th>Test Conditions (Fig. 4.1)</th>
<th>Recognition accuracy on ARR and AR database (%)</th>
<th>Baseline</th>
<th>Baseline Color</th>
<th>Final</th>
</tr>
</thead>
<tbody>
<tr>
<td>Session 1 images</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Smile,(b)</td>
<td></td>
<td>100</td>
<td>100</td>
<td>100</td>
</tr>
<tr>
<td>Anger,(c)</td>
<td></td>
<td>100</td>
<td>100</td>
<td>100</td>
</tr>
<tr>
<td>Scream,(d)</td>
<td></td>
<td>90</td>
<td>94</td>
<td>95</td>
</tr>
<tr>
<td>Left light,(e)</td>
<td></td>
<td>99</td>
<td>99</td>
<td>100</td>
</tr>
<tr>
<td>Right light,(f)</td>
<td></td>
<td>98</td>
<td>98</td>
<td>99</td>
</tr>
<tr>
<td>Both lights,(g)</td>
<td></td>
<td>81</td>
<td>87</td>
<td>99</td>
</tr>
<tr>
<td>Eye occlusion,(h)</td>
<td></td>
<td>100</td>
<td>100</td>
<td>100</td>
</tr>
<tr>
<td>Eye occlusion,left light,(i)</td>
<td></td>
<td>92</td>
<td>91</td>
<td>95</td>
</tr>
<tr>
<td>Eye occlusion,right light,(j)</td>
<td></td>
<td>91</td>
<td>94</td>
<td>93</td>
</tr>
<tr>
<td>Mouth occlusion,(k)</td>
<td></td>
<td>98</td>
<td>97</td>
<td>97</td>
</tr>
<tr>
<td>Mouth occlusion,left light,(l)</td>
<td></td>
<td>93</td>
<td>93</td>
<td>95</td>
</tr>
<tr>
<td>Mouth occlusion,right light,(m)</td>
<td></td>
<td>78</td>
<td>84</td>
<td>90</td>
</tr>
<tr>
<td>Session 2 images</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Neutral,(n)</td>
<td></td>
<td>97</td>
<td>97</td>
<td>98</td>
</tr>
<tr>
<td>Smile,(o)</td>
<td></td>
<td>87</td>
<td>88</td>
<td>90</td>
</tr>
<tr>
<td>Anger,(p)</td>
<td></td>
<td>94</td>
<td>92</td>
<td>95</td>
</tr>
<tr>
<td>Scream,(q)</td>
<td></td>
<td>57</td>
<td>58</td>
<td>57</td>
</tr>
<tr>
<td>Left light,(r)</td>
<td></td>
<td>92</td>
<td>94</td>
<td>97</td>
</tr>
<tr>
<td>Right light,(s)</td>
<td></td>
<td>89</td>
<td>87</td>
<td>92</td>
</tr>
<tr>
<td>Both lights,(t)</td>
<td></td>
<td>56</td>
<td>67</td>
<td>80</td>
</tr>
<tr>
<td>Eye occlusion,(u)</td>
<td></td>
<td>83</td>
<td>80</td>
<td>80</td>
</tr>
<tr>
<td>Eye occlusion,left light,(x)</td>
<td></td>
<td>78</td>
<td>81</td>
<td>80</td>
</tr>
<tr>
<td>Eye occlusion,right light,(v)</td>
<td></td>
<td>70</td>
<td>72</td>
<td>69</td>
</tr>
<tr>
<td>Mouth occlusion,(w)</td>
<td></td>
<td>80</td>
<td>79</td>
<td>82</td>
</tr>
<tr>
<td>Mouth occlusion,left light,(y)</td>
<td></td>
<td>64</td>
<td>64</td>
<td>78</td>
</tr>
<tr>
<td>Mouth occlusion, right light,(z)</td>
<td></td>
<td>54</td>
<td>58</td>
<td>65</td>
</tr>
<tr>
<td>Overall accuracy</td>
<td></td>
<td>84.8</td>
<td>86.2</td>
<td>89.0</td>
</tr>
</tbody>
</table>

*a* Image resolution is 66 × 48 pixels. The window size of the standard deviation filter is 3 × 3 pixels. The mean normalization window size is 33 × 24 pixels. The normalized distance n2 shown in Table 4.4 is used for the classifier. The value of θ is set to 0.25.

*b* Image resolution is 160 × 120 pixels. The window size of the standard deviation filter is 7 × 5 pixels. The mean normalization window size is 80 × 60 pixels. The normalized distance n2 shown in Table 4.4 is used for the classifier. The value of θ is set to 0.25.
### 4. Analysis of LBDS Algorithm

**Table 4.10: Comparison of the LBDS Algorithm with Other Algorithms (Single Training Sample per Person Problem)**

<table>
<thead>
<tr>
<th>Index (Fig. 4.1)</th>
<th>LBDS Baseline</th>
<th>LBDS Final</th>
<th>PCA</th>
<th>PCA²</th>
<th>SVD-PCA</th>
<th>LPS [28]</th>
<th>SOM-face [25]</th>
<th>SIS [24]</th>
<th>SIS$_{nondiv}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>(b-d)</td>
<td>97</td>
<td>99</td>
<td>87</td>
<td>87</td>
<td>87</td>
<td>83</td>
<td>95</td>
<td>99</td>
<td>85</td>
</tr>
<tr>
<td>(e-g)</td>
<td>93</td>
<td>100</td>
<td>65</td>
<td>77</td>
<td>45</td>
<td>N/A</td>
<td>N/A</td>
<td>100</td>
<td>99</td>
</tr>
<tr>
<td>(h)</td>
<td>100</td>
<td>100</td>
<td>56</td>
<td>48</td>
<td>44</td>
<td>80</td>
<td>97</td>
<td>99</td>
<td>87</td>
</tr>
<tr>
<td>(i-j)</td>
<td>92</td>
<td>98</td>
<td>22</td>
<td>17</td>
<td>16</td>
<td>N/A</td>
<td>N/A</td>
<td>96</td>
<td>60</td>
</tr>
<tr>
<td>(k)</td>
<td>98</td>
<td>99</td>
<td>13</td>
<td>10</td>
<td>7</td>
<td>82</td>
<td>95</td>
<td>98</td>
<td>89</td>
</tr>
<tr>
<td>(l-m)</td>
<td>86</td>
<td>97</td>
<td>8</td>
<td>6</td>
<td>6</td>
<td>N/A</td>
<td>N/A</td>
<td>97</td>
<td>77</td>
</tr>
<tr>
<td>(n)</td>
<td>97</td>
<td>100</td>
<td>81</td>
<td>80</td>
<td>80</td>
<td>N/A</td>
<td>100</td>
<td>100</td>
<td>95</td>
</tr>
<tr>
<td>(o-q)</td>
<td>79</td>
<td>87</td>
<td>63</td>
<td>62</td>
<td>60</td>
<td>76</td>
<td>81</td>
<td>86</td>
<td>60</td>
</tr>
<tr>
<td>(r-t)</td>
<td>79</td>
<td>97</td>
<td>34</td>
<td>27</td>
<td>24</td>
<td>N/A</td>
<td>N/A</td>
<td>99</td>
<td>89</td>
</tr>
<tr>
<td>(u)</td>
<td>83</td>
<td>97</td>
<td>29</td>
<td>26</td>
<td>25</td>
<td>54</td>
<td>60</td>
<td>96</td>
<td>61</td>
</tr>
<tr>
<td>(v-w)</td>
<td>74</td>
<td>83</td>
<td>14</td>
<td>10</td>
<td>9</td>
<td>N/A</td>
<td>N/A</td>
<td>82</td>
<td>36</td>
</tr>
<tr>
<td>(x)</td>
<td>80</td>
<td>92</td>
<td>8</td>
<td>6</td>
<td>6</td>
<td>48</td>
<td>53</td>
<td>90</td>
<td>78</td>
</tr>
<tr>
<td>(y-z)</td>
<td>59</td>
<td>83</td>
<td>5</td>
<td>3</td>
<td>2</td>
<td>N/A</td>
<td>N/A</td>
<td>84</td>
<td>63</td>
</tr>
</tbody>
</table>

| Average          | 84.8          | 94.3       | 37.3| 35.3 | 31.6    | 94.1     | 76.3         |

*Results not available from the literature.*

From session one (Fig. 4.1) are used to form the gallery set, whereas a total of 2500 images from both session one and session two form the test set. The performances of three different versions of the LBDS algorithm are shown in Table 4.9. The LBDS baseline algorithm and the LBDS baseline color algorithm use gray scale images and color images, respectively, and do not employ any method for compensating localization errors. The LBDS final algorithm uses color images and compensates for localization errors by generating synthetic gallery images by randomizing the eye coordinates. Table 4.9 shows the results of comparing the three versions of the algorithm using both AR and ARR databases.

#### 4.3.4.1 Faces with Occlusion

Occlusions in any form reduce the available identity information. This results in a loss of useful spatial change features required for recognition. Further, occlusions caused by objects like scarfs and glasses can trick the feature extractors. These effects reduces the recognition performance of most face recognition algorithms considerably.
4. Analysis of LBDS Algorithm

The LBDS algorithm treats occlusions in the same way as noise. By applying binary decisions, the mismatch caused by an occlusion is reduced. It can be observed from Table 4.9 that the proposed method is almost invariant to the effects of occlusions provided in the AR database, which makes it useful for practical applications.

4.3.4.2 Faces with Illumination

Illumination modifies the image pixel intensities and hence modifies the texture information. This results in a loss of information. However, as distinct from faces with occlusion, spatial intensity change features will be preserved if the morphology of the face images is not destroyed and if the light variation is uniform. It is observed that gray images result in loss of this information and might require efficient coding schemes to estimate the missing features. A simple way to get around this problem is using color images that preserve features along its individual channels. It is observed that recognition improves by using color images. This is evident from the results obtained when using images with varying directional illumination and those with the combined effects of occlusion and illumination (Table 4.9).

4.3.4.3 Faces with Expressions

Expressions in images can also result in loss of identity information. However, major features are still preserved which makes this a relatively easier problem. Apart from the screaming expression in session two, where the face in the image is spatially disoriented, the recognition performance remains robust (Table 4.9).

4.3.4.4 Comparison with Other Algorithms

The ARR database is used for comparison with other algorithms to ensure a fairness in comparison of recognition performance. It should be noted that the
use of AR database rather than ARR can result in slightly better recognition rates due to higher image resolution, as shown in Table 4.9. As can be seen from Table 5.5, the LBDS algorithm considerably outperforms the major PCA based algorithms. This difference in performance arises mainly because the PCA methods use an NN classifier and do not work with face images containing occlusions. The failure of PCA can be attributed to the inability of PCA to address the effects of occlusion in face images. LPS was one of the first successful methods that attempted to solve the problem caused by face images with occlusions. However, their recognition performance was poor with increased variability in the images. SOM-face based on neural network shows good accuracies on specific conditions but is computationally expensive and requires manual removal of the occlusions from the images. A contemporary algorithm that shows very similar recognition accuracies to the LBDS algorithm is the SIS algorithm [24]. It is observed that the performance of SIS can be attributed to (1) local decisions made at block level and (2) various filtering schemes that try to remove the effects of illumination on images. However, the drawbacks of such a scheme are (1) the inability to make the local decisions at pixels (the smallest possible block), (2) the need of many different filter combinations, and (3) much higher computational complexity. The superiority of local binary decisions used by LBDS can be asserted from Fig. 4.10, which shows that increasing the number of decision making blocks (maximum is the number of pixels), improves the recognition performance. The ability of LBDS to make decisions at pixel level implies that it is able to utilize the maximum performance offered by local decisions. Further, LBDS exhibits much lower computational complexity and storage requirements.

4.3.5 Experiments on FERET

A study of recognition performance with a large gallery size is very important to check the robustness of an algorithm for practical applications. The gallery
set Fa with 1,196 images is used to generate 15 synthetic images per person by perturbations of the eye coordinates. The test-set images are then compared with all the images in the gallery and the one with the maximum similarity score is taken as the best match.

As can be seen from Table 5.7, the LBDS method performs better than other reported methods. The CSU face evaluation system [129] was employed to benchmark the recognition performance of other algorithms shown in Table 5.7. The results of the SIS method are directly adopted from the literature. To make a fair comparison, the same set of images are selected for the LBDS algorithm on the tests performed with the CSU face evaluation system.

### 4.3.5.1 Faces with Expression

The test set Fb contains images with various random expressions. Random expressions over a large gallery represent a very practical scenario for face recognition applications. In addition, the recognition accuracy decreases with increased gallery size. The tests on the Fb set show that the LBDS final algorithm outperforms the algorithms in the CSU evaluation system.

### 4.3.5.2 Faces with Illumination

The test set Fc contains images with the effect of illumination. Illumination can have significant impact on the recognition performance of an algorithm. The LBDS final algorithm shows high recognition performance in comparison to the standard algorithms used in the CSU evaluation system. This is due to the invariant nature of the spatial change features with respect to illumination and the superior performance due to the local binary decisions.
4. Analysis of LBDS Algorithm

Table 4.11: Recognition Results on FERET Database

<table>
<thead>
<tr>
<th>Methods</th>
<th>Recognition accuracy on FERET database (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>LBDS final</td>
<td>Fa-Fb</td>
</tr>
<tr>
<td>SIS [24]</td>
<td>91</td>
</tr>
<tr>
<td>PCA, MahCosine [129]</td>
<td>85</td>
</tr>
<tr>
<td>PCA_Euclidean [129]</td>
<td>74</td>
</tr>
<tr>
<td>Bayesian,MAP [129]</td>
<td>82</td>
</tr>
<tr>
<td>EBGM Standard [129]</td>
<td>88</td>
</tr>
<tr>
<td>LDA_Euclidean [129]</td>
<td>61</td>
</tr>
</tbody>
</table>

4.3.5.3 Faces with Aging

The test set Dup1 and Dup2 contains images with the effect of aging. Aging of the faces is a difficult problem in face recognition with a single training sample per person. The increased time gap between the images in the gallery set and the test set results in decreased recognition performance. Aging causes natural and artificial changes to face images which can be seen as random noise that cannot be separated from the signal. We can observe that the LBDS final algorithm outperform the benchmark algorithms in the CSU face evaluation system, although it is not highly robust against such drastic variations.

4.4 Conclusions

In this chapter a detailed analysis of the LBDS algorithm is presented and a comparison with other relevant face recognition algorithms. The analysed algorithm outperforms the algorithms that are either known to show high performance with limited training images per person or that are widely used to address this type of problem. The main attributes that contribute to the high recognition accuracy of the LBDS algorithm are (1) conceptual application of local binary decisions and (2) the use of a modular design to form a system that helps to improve the recognition performance of the base algorithm. The local binary
decisions on spatial change features at the pixel level are the main contributing factor for the high recognition accuracies. It is observed that local decisions, when used by any algorithm, increase its recognition performance. The use of local binary decisions enables recognition under difficult conditions such as images with occlusions. This advantage of local binary decisions arises from their ability to remove noise. Variability in face images reduces the intra-class similarity, however, the application of binary decisions results in a more effective use of this similarity. The removal of false similarity between the features and identifying the most similar features using a global threshold results in tolerance to natural variability. It was found that normalization of feature vectors and similarity measures in combination with local binary decisions helps to increase the recognition performance. Further, normalization increases the range of global threshold values that can be used. Resolution of the features is another contributing factor towards attaining increased performance. Increased resolution of features and utilizing all these features for making local binary decisions results in increased recognition performance. The use of color images can further increase the recognition performance. The color images are incorporated into the algorithm using the idea of modularity. Each channel in the image is treated as an independent module and combined into a system. Localization errors in the detected faces are yet another important contributing factor that affects the recognition accuracy. This is again handled through the modular design, by generating synthetic images (or rather shifted features). A higher recognition accuracy is observed when using schemes that compensate for the localisation errors.

The SIS method, which shows one of the best reported recognition accuracies, exhibits a high computational complexity because it requires various combination of spatial filters for its working and cannot utilize the maximum possible image resolution (for decisions on features at the pixel level). This inability to
employ local decisions at individual pixels makes SIS inferior to LBDS. SOM-face method based on neural networks shows good accuracies on specific conditions but is computationally expensive and uses manual removal of occlusions from the images. Other methods are mathematically complex and/or do not show comparable recognition performance. The LBDS algorithm is a generic algorithm that can find applications in various visual pattern recognition tasks.
Chapter 5

Enhanced LBDS Algorithm

5.1 Introduction

In this chapter, the use of spatial filtering operations to improve the performance of the LBDS algorithm is shown and this improved algorithm is named Enhanced Local Binary Decisions on Similarity (ELBDS) algorithm. In this algorithm, spatial filtering is integrated into the feature extraction stage along with the spatial change detector which results in the extraction of more useful information that is required for recognition. Also local binary decisions are made on average similarity measure formed as a result of comparison between different feature vectors that are extracted. This chapter shows that the addition of these components improves the recognition performance considerably and show superior performance against other major methods. Throughout this chapter the experiments are carried out using single gallery per person on all the database used for simulation. In fact, this is one of the most difficult technical problem in face recognition due to: (1) limitation in training data, and (2) the high variability between the images in gallery set and test set [21]. And, this specific face recognition problem is important to address due to: (1) its application in wide range of biometric and law enforcement methods, and (2) could provide highly
5. Enhanced LBDS Algorithm

accurate face recognition with the addition of a greater number of gallery images per person in real-time applications.

This chapter is organised into the following sections, (1) Section 5.2 shows a brief overview of the ELBDS algorithm, (2) the analysis describing the importance of spatial filtering along with a study on the rationale of using different average similarity measures is shown in Section 5.3, (3) Section 5.4 shows the comparison of the algorithm with other well known methods reported and with a few different databases, and (4) the final section includes the conclusion and discussions relevant to the aims of this chapter.

5.2 ELBDS Algorithm

The two main parts in ELBDS algorithm are: (1) a feature extractor that extracts texture variation using different spatial filters followed by normalized spatial change features calculation, and (2) an ELBDS classifier based on the local binary decisions on a normalized similarity measure. Texture variation can be extracted by local spatial filters such as the Gradient filters. This is followed by finding spatial change features by using a local standard deviation filter. The normalization of spatial change features is performed by local mean normalization on each feature. Once the features are extracted from the two images under comparison, the similarity is calculated by doing a feature by feature comparison using a normalized similarity measure $\delta$. Making local binary decisions on this similarity measure forms a binary decision vector. Summing the value of all the elements in this decision vector forms the similarity score $S_g^*$. Finally, ranking all the similarity scores [which results due to the comparison of a test image with many gallery images] completes the algorithm. Figure 5.1 shows a graphical overview of the ELBDS feature extractor and classifier that form the basic blocks in the ELBDS algorithm.
Figure 5.1: The figure illustrates the two basic blocks of the ELBDS algorithm using face images from AR database [1,2]. The feature extraction stage consists of spatial filtering followed by spatial change detection. The six spatial filters shown in Table 5.1 are used for illustration. The ELBDS classifier consists of normalized similarity measure calculation following by finding local binary decisions and similarity score calculation. These blocks can be replicated to account for color images and to compensate for localization error as described in Section 5.2.2.3 and Section 5.2.2.4.
5.2.1 Feature Extractor

Each pixel intensity in the raw image is denoted as $I(i,j)$, where $(i,j)$ represents the location of a pixel in the image space. Spatial filtering is applied on such an image to form the corresponding filtered image intensities $Y^*(i,j)$. This is used to calculate the normalized spatial change features denoted as $x(i,j)$.

5.2.1.1 Spatial Filtering for Information Extraction

Spatial filtering is a preprocessing step in the presented ELBDS algorithm which helps to maximize the utilization of identity information in an image. Filtering is done using a moving rectangular window $w_f^{(p)}$ of size $n \times m$ over the image $I$ of size $N \times M$. This process represents a linear convolution and is expressed as:

$$Y^{(p,*)}(i,j) = \sum_{s=-a0}^{a0} \sum_{t=-b0}^{b0} w_f^{(p)}(s,t)I(s+i,t+j) \quad (5.1)$$

where, $a0 = (n - 1)/2$, $b0 = (m - 1)/2$, $p = 1, \ldots, P$ and $P$ is the maximum number of the filters applied. There are many ways to design spatial change pre-processing filters. Some of the examples we use are (1) Gradient based filters and (2) Gabor based filters.

**Gradient based filters** A symmetric weight matrix is used for forming the moving filter window [see Table 5.1]. The operations implemented using the filter widow are: (1) image and image averaging [see index (ta) and (tb) listed in Table 5.1], (2) combinations of vertical, horizontal and diagonal difference in the image [see index (tc) and (td) listed in Table 5.1], and (3) edge variations [see index (te) and (tf) listed in Table 5.1]. In the remaining sections in this thesis, we use the terminology ELBDS filters to represent the Gradient filters described in Table 5.1.
5. Enhanced LBDS Algorithm

**Gabor based filters**  The V1-like features generated from Gabor filters are yet another way to detect different types of spatial variations in the images. The advantage of Gabor filters for feature extraction in face recognition was evident through the works of [130]. These suggest that like the Gradient filters Gabor filters can be used for preprocessing the images. Formally, Gabor filters are defined as:

\[
\psi_{\mu,\nu}(z) = \frac{k_{\mu,\nu}^2}{\sigma^2} e^{-\|k_{\mu,\nu}^2\|z^2/2\sigma^2} [e^{ik_{\mu,\nu}z} - e^{-\sigma^2/2}] \tag{5.2}
\]

where \(\mu\) defines the orientation, \(\nu\) defines the scale of the Gabor filters, \(k_{\mu,\nu} = \frac{k_{\text{max}}}{\lambda} e^{i\frac{\pi}{8}}\), \(\lambda\) is the spacing between the filters in frequency domain and \(\|\cdot\|\) denotes the norm operator. The phase information from these filters are not considered, and only its magnitude explored. For the experiments, we set the value of parameters as follows: \(\lambda = \sqrt{2}, \sigma = 2\pi\) and \(k_{\text{max}} = \pi/2\). Further by considering five scales \(\nu \in \{0, \ldots, 4\}\) and eight orientations \(\mu \in \{0, \ldots, 7\}\) which on convolution using Eq. 5.1 results in 40 filters.

### 5.2.1.2 Spatial Change Features

Spatial change features of an image can be detected by calculating the local standard deviation across a window of pixels. A local standard deviation filter is employed to extract the spatial change features by the following operation:

\[
\sigma^{(p,i)}(i,j) = \sqrt{\frac{1}{mn} \sum_{s=-a}^{a} \sum_{t=-b}^{b} [Y^{(p,i)}(i+s,j+t) - \overline{Y^{(p,i)}(i,j)}]^2} \tag{5.3}
\]

where \(a = (m-1)/2\) and \(b = (n-1)/2\). The local mean \(\overline{Y^{(p,i)}(i,j)}\) used in (5.3) is calculated by the following equation:

\[
\overline{Y^{(p,i)}(i,j)} = \frac{1}{mn} \sum_{s=-a}^{a} \sum_{t=-b}^{b} Y^{(p,i)}(i+s,j+t) \tag{5.4}
\]
Table 5.1: Various Spatial filter operations used in the preprocessing stage of ELBDS algorithm

<table>
<thead>
<tr>
<th>Index</th>
<th>Filter name</th>
<th>Window weights</th>
</tr>
</thead>
</table>
| (ta)  | All-Pass                          | \[
|       |                                   | \begin{bmatrix} 0 & 0 & 0 \\
|       |                                   | 0 & 1 & 0 \\
|       |                                   | 0 & 0 & 0 \\
|       |                                   | 1 & 1 & 1 \end{bmatrix} |
| (tb)  | Average                           | \[
|       |                                   | \begin{bmatrix} \frac{1}{9} & 1 & 1 \\
|       |                                   | 1 & 1 & 1 \\
|       |                                   | 0 & -1 & 0 \end{bmatrix} |
| (tc)  | Horizontal-Vertical Difference    | \[
|       |                                   | \begin{bmatrix} -1 & 4 & -1 \\
|       |                                   | 0 & -1 & 0 \\
|       |                                   | -1 & 0 & -1 \end{bmatrix} |
| (td)  | Diagonal Difference               | \[
|       |                                   | \begin{bmatrix} 0 & 4 & 0 \\
|       |                                   | -1 & 0 & -1 \\
|       |                                   | 1 & 2 & 1 \end{bmatrix} |
| (te)  | Sobel edge operator               | \[
|       |                                   | \begin{bmatrix} 0 & 0 & 0 \\
|       |                                   | -1 & -2 & -1 \\
|       |                                   | -1 & 0 & 1 \end{bmatrix} |
| (tf)  | Sobel edge operator               | \[
|       |                                   | \begin{bmatrix} -2 & 0 & 2 \\
|       |                                   | -1 & 0 & 1 \end{bmatrix} |
5. Enhanced LBDS Algorithm

5.2.1.3 Local Mean Normalization

Normalization of spatial change features is an important operation that helps to improve the recognition performance of ELBDS algorithm. The normalized spatial change features is calculated using the following equation:

$$x^{(p,*)}(i, j) = \frac{\sigma^{(p,*)}(i, j)}{\sigma^{(p,*)}(i, j)}$$  \hspace{1cm} (5.5)

where, \(\sigma^{(p,*)}(i, j)\) represents the local mean of spatial change features in the window of features of size \(k \times l\) pixels. The value of \(\sigma^{(p,*)}(i, j)\) is calculated using the following equation:

$$\sigma^{(p,*)}(i, j) = \frac{1}{k l} \sum_{s=-a1}^{a1} \sum_{t=-b1}^{b1} \sigma^{(p,*)}(i + s, j + t)$$  \hspace{1cm} (5.6)

where, \(a1 = (k - 1)/2\) and \(b1 = (l - 1)/2\).

5.2.2 ELBDS Classifier

In a generic face recognition task a test image \(Y_t^{(p)}\) is compared with a finite number of gallery images \(Y_g^{(p,k)}\). For each comparison a similarity score is determined and the one having the highest score is determined to be the best match which determines the class of the test image. The classification process starts with finding the similarity measure vectors \(\delta^{(p,*)}\), where \(k\) represents the index of the person in the gallery, \(k = 1 \ldots K\). These similarity measure vectors \(\delta^{(p,*)}\) are a result of comparison between normalized spatial change features \(x_t^{(p)}\) formed from a test image \(Y_t\) and a set of normalized spatial change feature vectors \(x_g^{(p,k)}\) formed from test image \(Y_g^{(p,k)}\).

5.2.2.1 Normalized Similarity and Binary Decisions

A normalized similarity measure is obtained by using an absolute difference measure. Smaller the value of this difference more will be the similarity of the
Table 5.2: Spatial filtering window weights of various methods used for extracting various spatial information from an image

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>0 1 0</td>
<td>0 1 0</td>
<td>1 -2 1</td>
<td>2 0 -1</td>
<td>5 0 -3</td>
</tr>
<tr>
<td>0 0 0</td>
<td>0 0 0</td>
<td>-1 -1 -1</td>
<td>1 0 -1</td>
<td>5 -3 -3</td>
</tr>
<tr>
<td>1 1 1</td>
<td>1 1 1</td>
<td>1 1 1</td>
<td>2 1 0</td>
<td>5 5 -3</td>
</tr>
<tr>
<td>1 1 1</td>
<td>1 1 1</td>
<td>1 -2 -1</td>
<td>1 0 -1</td>
<td>5 0 -3</td>
</tr>
<tr>
<td>1 1 1</td>
<td>1 1 1</td>
<td>-1 -1 -1</td>
<td>1 0 -1</td>
<td>5 -3 -3</td>
</tr>
<tr>
<td>0 -1 0</td>
<td>-1 -1 -1</td>
<td>1 1 -1</td>
<td>1 2 1</td>
<td>5 5 5</td>
</tr>
<tr>
<td>-1 4 -1</td>
<td>-1 8 -1</td>
<td>1 -2 -1</td>
<td>0 0 0</td>
<td>-3 0 -3</td>
</tr>
<tr>
<td>0 -1 0</td>
<td>-1 -1 -1</td>
<td>1 1 -1</td>
<td>-1 2 -1</td>
<td>-3 -3 -3</td>
</tr>
<tr>
<td>-1 0 -1</td>
<td>-1 -1 -1</td>
<td>1 -1 -1</td>
<td>0 1 2</td>
<td>-3 5 5</td>
</tr>
<tr>
<td>0 4 0</td>
<td>-1 9 -1</td>
<td>1 -2 -1</td>
<td>-1 0 1</td>
<td>-3 0 5</td>
</tr>
<tr>
<td>-1 0 -1</td>
<td>-1 -1 -1</td>
<td>1 1 -1</td>
<td>-2 -1 0</td>
<td>-3 -3 -3</td>
</tr>
<tr>
<td>1 2 1</td>
<td>1 0 0</td>
<td>-1 -1 -1</td>
<td>-1 0 1</td>
<td>-3 -3 -3</td>
</tr>
<tr>
<td>0 0 0</td>
<td>0 -1 0</td>
<td>1 -2 1</td>
<td>-2 0 2</td>
<td>-3 0 5</td>
</tr>
<tr>
<td>-1 -2 -1</td>
<td>0 0 0</td>
<td>1 1 1</td>
<td>-1 0 1</td>
<td>-3 -3 -5</td>
</tr>
<tr>
<td>-1 0 1</td>
<td>1 0 0</td>
<td>-2 -1 1</td>
<td>-1 0 1</td>
<td>-3 0 5</td>
</tr>
<tr>
<td>-2 0 2</td>
<td>0 2 0</td>
<td>-2 -1 1</td>
<td>-1 0 1</td>
<td>-3 0 5</td>
</tr>
<tr>
<td>-1 0 1</td>
<td>0 0 -1</td>
<td>1 1 1</td>
<td>0 1 2</td>
<td>-3 -3 5</td>
</tr>
<tr>
<td>1 2 1</td>
<td>1 1 1</td>
<td>-1 -2 1</td>
<td>-1 0 1</td>
<td>-3 -3 -3</td>
</tr>
<tr>
<td>0 0 0</td>
<td>-1 -2 1</td>
<td>1 2 1</td>
<td>5 5 5</td>
<td></td>
</tr>
<tr>
<td>-1 -2 -1</td>
<td>-1 1 1</td>
<td>0 -1 -2</td>
<td>-3 -3 -3</td>
<td></td>
</tr>
<tr>
<td>4 8 4</td>
<td>1 1 1</td>
<td>0 -1 -2</td>
<td>5 0 -3</td>
<td></td>
</tr>
</tbody>
</table>
comparing features. The feature by feature comparison is achieved by using the following equation:

\[
\delta^{(p,d)}(i, j) = \frac{|x_g^{(p,k)}(i, j) - x_t^{(p,*)}(i, j)|}{\min(x_g^{(p,k)}(i, j), x_t^{*}(i, j))}
\] (5.7)

This is followed by element-wise averaging of the normalized differences obtained along different filters. This is expressed as:

\[
\hat{\delta}^{(k)}(i, j) = \frac{1}{P} \sum_{p=1}^{P} \delta^{(p,k)}(i, j)
\] (5.8)

Local binary decisions are made on \(\hat{\delta}^{(k)}(i, j)\) by applying a global threshold \(\theta\). This operation results in a binary decisions vector \(B^{(k)}\) represented as:

\[
B^{(k)}(i, j) = \begin{cases} 
1 & \hat{\delta}(i, j) < \theta \\
0 & \hat{\delta}(i, j) \geq \theta 
\end{cases}
\] (5.9)

### 5.2.2.2 Global Similarity and Ranking

The binary decisions vector represents decisions on pixel wise feature comparisons between a test and a gallery image. Summing up all the decisions results in a global similarity score \(S_g^{(k)}\):

\[
S_g^{(k)} = \sum_{i=1}^{N} \sum_{j=1}^{M} B^{(k)}(i, j)
\] (5.10)

where \(B\) consists of \(N \times M\) binary decisions. The comparison of a test image with \(k\) gallery images results in \(k\) similarity scores.

### 5.2.2.3 Use of Color Images

ELBDS algorithm can also be used with color images to use the additional information available in it. A color image is formed of three channels namely, red, green and blue. Each of the red, green and blue channels in the image results
in a feature vector denoted as \( x^{(s)}_r \), \( x^{(s)}_g \) and \( x^{(s)}_b \). A one to one comparison of the channels of a test image having feature vectors \( x^{(p)_r}_t \), \( x^{(p)_g}_t \) and \( x^{(p)_b}_t \) with a gallery image having feature vectors \( x^{(p,k)_r}_g \), \( x^{(p,k)_g}_g \) and \( x^{(p,k)_b}_g \) respectively, results in three similarity scores \( S^{(k)}_r \), \( S^{(k)}_g \) and \( S^{(k)}_b \). The final similarity score \( S^{(k)}_g \) is calculated using the following equation:

\[
S^{(k)}_g = \max \left[ S^{(k)r}_{g}, S^{(k)g}_{g}, S^{(k)b}_{g} \right] \tag{5.11}
\]

### 5.2.2.4 Localization Error Compensation

Irrespective of how faces are detected and aligned in the images localization errors occur that effect overall recognition performance of face recognition system. In most practical situations these errors are difficult to be quantified and are random in nature. To deal with this problem, we use one of the simple ways of compensating localization error is by the perturbation of test image \( x_t \) to generate a set of \( T \) test images \( x^{(t1)}_t \), where \( t1 = 1, 2, \ldots, T \). The comparison of a set of test images \( x^{(t1)}_t \) with a gallery image \( x^{(k)}_g \) results in \( T \) similarity scores \( S^{(t1)}_g \). The maximum among these similarity scores represents the final similarity score \( S^{(k)}_g \) for the comparison between the test image \( x_t \) and and the gallery image \( x^{(k)}_g \).

\[
S^{(k)}_g = \max \left[ S^{(1)}_g, S^{(2)}_g, \ldots, S^{(t1)}_g, \ldots, S^{(T)}_g, S^{(T)}_g \right] \tag{5.12}
\]

A second method to compensate for the localization error is by the perturbation of gallery image \( x^{(k)}_g \) to generate a set of \( G \) test images \( x^{(g1)}_g \), where \( g1 = 1, 2, \ldots, G \). The comparison of a test images \( x^{(d)}_t \) with a set of generated gallery images \( x^{(g1)}_g \) of a person results in \( G \) similarity scores \( S^{(g1)}_g \). The maximum among these similarity scores represents the final similarity score \( S^{(d)}_g \) for the comparison between the test image \( x_t \) and and the gallery image \( x^{(g1)}_g \).

\[
S^{(k)}_g = \max \left[ S^{(1)}_g, S^{(2)}_g, \ldots, S^{(g1)}_g, \ldots, S^{(G)}_g, S^{(G)}_g \right] \tag{5.13}
\]
5. Enhanced LBDS Algorithm

5.2.2.5 Ranking and Best Match

Ranking the \( K \) final similarity scores obtained as result of comparisons on \( K \) classes and selecting the one with maximum similarity score determines the class of the test image. This operation is summarized as:

\[ k^* = \arg \max_k S_k^{(k)} \]  \hfill (5.14)

5.3 Analysis of Spatial Filtering and Classifier

5.3.1 Database and Setup

The AR database \([1, 2]\) contains images of 126 persons taken in two different sessions. In each session photos were taken in 13 different conditions [see Fig. 4.1]. They are: (1) neutral, (2) expressions [anger, scream, and smile], (3) illumination [right, left and both sides], and (4) occlusions [eye occlusion, eye occlusion with left illumination, eye occlusion with right illumination, mouth occlusion, mouth occlusion with left illumination and mouth occlusion with right illumination].

In this chapter, single gallery image per person is used to form the gallery set. The image with neutral expression taken on first session is used for this [see Fig. 4.1]. The remaining 25 images of the same person that has different natural variability to it is used as test images [see Fig. 4.1]. Specifically, the tests are done on 100 persons selected randomly from the AR database. Further, in this selection of 100 persons, 50 are men and 50 are women. All the selected images are cropped to a image size of 160 \( \times \) 120 pixels based on the eye-coordinates of the face located manually.

For the results reported using AR database following are the parameters which was determined empirically: (1) global threshold \( \theta \) is set to 0.25, (2) input image size is resized to 60 \( \times \) 60 pixels irrespective of actual image dimension, (3)
Table 5.3: Comparison of information extraction using ELBDS algorithm

<table>
<thead>
<tr>
<th>Method</th>
<th>Recognition accuracy (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Multiple filters</strong></td>
<td></td>
</tr>
<tr>
<td>ELBDS filters(^c)</td>
<td>92.9</td>
</tr>
<tr>
<td>SIS filters</td>
<td>92.6</td>
</tr>
<tr>
<td>Gabor filters</td>
<td>92.9</td>
</tr>
<tr>
<td>Prewitt filters</td>
<td>89.2</td>
</tr>
<tr>
<td>Sobel filters</td>
<td>89.4</td>
</tr>
<tr>
<td>Krisch filters</td>
<td>89.3</td>
</tr>
<tr>
<td><strong>Single filter</strong></td>
<td></td>
</tr>
<tr>
<td>All-Pass (ta)</td>
<td>88.7</td>
</tr>
<tr>
<td>Average (tb)</td>
<td>85.8</td>
</tr>
<tr>
<td>Horizontal-Vertical difference (tc)</td>
<td>83.0</td>
</tr>
<tr>
<td>Diagonal difference (td)</td>
<td>84.0</td>
</tr>
<tr>
<td>Sobel edge operator (te)</td>
<td>81.0</td>
</tr>
<tr>
<td>Sobel edge operator (tf)</td>
<td>81.0</td>
</tr>
</tbody>
</table>

\(^a\) These filters has been described in Table 5.2. The ELBDS algorithm with the same global parameters are used with all the filters.

\(^b\) The results are shown when no perturbations are applied to gallery or test images.

\(^c\) See Table 5.1

spatial change window size is kept at 3 × 3 pixels, (4) local normalization window size is kept at 30 × 30 pixels, and (5) the maximum number of perturbations applied \(T\) is set at 30, that includes \(\pm 5\) vertical shifts, \(\pm 5\) horizontal shifts and \(\pm 5\) diagonal shifts.

### 5.3.2 Preprocessing Spatial Filters

In general, spatial filtering methods are used for image enhancement and for the detection of edges in an image. Edge detection techniques detects the inter-pixel
Figure 5.2: Graphical illustration showing the dependence of recognition performance with the use of number of ELBDS preprocessing spatial filters when using color and gray versions of images in AR database.
spatial variation that occur locally in an image. These spatial variations are extracted as a measure of gradient between a pixel and its surrounding pixels using a moving filter window [or in other words convolution, see Eq. (5.1)]. Some of the well known methods for extracting the gradient are the Prewitt, Sobel and Krisch filter windows [131]. One another recently proposed method that uses combination of filters for extracting features and that has been successfully applied for face recognition is the SIS method [24]. The coefficients [or weights] of these filter windows are listed in the Table 5.2.

To study the effect of using filtering methods listed in Table 5.2, ELBDS algorithm is used with a constant set of parameters described in Section 5.3.1. The result of this analysis using AR database is shown in Table 5.3. Clearly, the ELBDS filters show a better recognition performance and use fewer number of filters to achieve it. This further shows that the use of multiple preprocessing spatial filters results in increased recognition accuracy as opposed to when using only a single preprocessing spatial filter. It can be seen from Table 5.3 that there is a minimum 5% increase in the recognition accuracy when multiple preprocessing spatial filters are applied. Furthermore, this also means that each of the six ELBDS filters provide some additional information required for recognition as shown in Fig. 5.2.

Spatial change detection is also a spatial filtering method and is the major contributing element towards the recognition performance as seen from the results of All-Pass filter in Table 5.3. If spatial change features are not used and instead the filtered outputs of the ELBDS filters are used directly, results in a drop recognition accuracy from 92.8% to 10%. This shows the important role of spatial change features as an essential component of ELBDS algorithm in providing a high recognition performance and the role of other spatial filters in improving the recognition performance and robustness.
5. Enhanced LBDS Algorithm

5.3.3 Local Binary Decisions on Similarity Averages

Another important aspect is the way in which the filter outputs are combined. Apart from taking the average of the normalized difference as shown in Eq. (5.8), one other method is to take the geometric mean given by the following equation:

\[ \hat{\delta}^{(k)}(i, j) = \left( \prod_{p=1}^{P} \delta^{(p,k)}(i, j) \right)^{\frac{1}{P}} \]  

(5.15)

In both cases of averaging, shown by Eq. (5.8) and Eq. (5.15), the idea is to reinforce the available similarity. The information unique to the matched identity of the face will be available in each of the ELBDS filtered outputs. The repetition of this matching features is made use in the processes of averaging. The simulation using color images in AR database and without applying perturbations show a recognition accuracy of 91.9% when Eq. (5.15) is used, while using Eq. (5.8) results in a recognition accuracy of 92.9%. To ascertain the idea of matched repetition, if we assume that there is no repetition of the matched information, it is worthwhile to select the minimum distance among all the filtered outputs. This is given by:

\[ \hat{\delta}^{(k)}(i, j) = \min \left( \delta^{(1,k)}(i, j), \ldots, \delta^{(P,k)}(i, j) \right) \]  

(5.16)

Using the same conditions described for simulating Eq. (5.15), the recognition accuracy using Eq. (5.16) is 82%. This clearly shows that across the different ELBDS filters similarity repetition occur at the same spatial locations and justifies the use of averaging of similarity information. Furthermore, as a corollary to this, if all the similarity information repeats at the same or uniquely different spatial locations then the use of multiple spatial filters will not improve the recognition performance.
Table 5.4: Overall Recognition performance on AR database

<table>
<thead>
<tr>
<th>Image type</th>
<th>Preprocessing spatial filters used</th>
<th>No</th>
<th>Yes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gray</td>
<td></td>
<td>85</td>
<td>91</td>
</tr>
<tr>
<td>Color</td>
<td></td>
<td>89</td>
<td>93</td>
</tr>
</tbody>
</table>

Top rank recognition accuracy with localization error compensation (%)

<table>
<thead>
<tr>
<th>Image type</th>
<th>Preprocessing spatial filters used</th>
<th>No</th>
<th>Yes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gray</td>
<td></td>
<td>89</td>
<td>95</td>
</tr>
<tr>
<td>Color</td>
<td></td>
<td>94</td>
<td>97</td>
</tr>
</tbody>
</table>

5.4 Experimental Results and Comparisons

This section contains a comparative analysis and performance of the ELBDS algorithm when using single gallery image per person in the following databases: (1) AR database to show the robustness against different types of variability, (2) FERET database to benchmark the performance against a large gallery database, (3) YALE database to show the performance on a small gallery database with random variability, (4) Extended Yale (EYALE) database to show the performance on a wide range of illumination variation, and (5) CALTECH database to show small gallery performance on photos taken on random locations with different uncontrolled variations.

5.4.1 Experiments with AR Database

An analysis is performed on AR database to benchmark the overall recognition performance of the ELBDS algorithm under different conditions. For this, a
comparison is done when the ELBDS algorithm is applied with or without the use of color images, localization error compensation and spatial filtering. The results of this comparison are shown in Table 5.4. Clearly, it can be seen that availability of more texture information shows a considerable improvement in the recognition accuracy. The use of color images and preprocessing spatial filters are used to extract as much texture information as possible. Further, in the case of single training image per person problem, localization compensation improves the recognition performance of the system.

To compare the recognition performance of the ELBDS algorithm with other well known or best performing recognition algorithms, again use the parameter setting described in the Section 5.3.1. A brief comparison is performed against the following algorithms: PCA [10–12], SIS [24], SIS_{nondiv} [24], LPS [28], and SOM-face [25]. The results of the comparison is shown in Table 5.5. It can be seen that the ELBDS algorithm outperforms the other algorithms considerably with respect to the recognition accuracy. These high recognition accuracies of the ELBDS algorithm are attributed to the combined use of feature extraction and local binary decisions. The use of preprocessing spatial filters, and color images increases the amount of information available for recognition. Localization error compensation reduces the effect of feature mismatch during the comparisons and provides a better chance to find a best match. Further, spatial change detection reduces the effects of dc offsets in the image, reduces the variability of illumination and provides an easy way to extract the spatial variability. Local binary decisions removes the noisy similarity values during comparison, which helps in improving the recognition performance across different natural variabilities resulting as a result of occlusion or noise.
Table 5.5: Comparison of the ELBDS algorithm with other algorithms using single gallery image person in the AR database.

<table>
<thead>
<tr>
<th>Test conditions</th>
<th>Recognition accuracy on AR database (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>ELBDS&lt;sup&gt;a&lt;/sup&gt;</td>
</tr>
<tr>
<td>Session 1 images</td>
<td>100</td>
</tr>
<tr>
<td></td>
<td>100</td>
</tr>
<tr>
<td></td>
<td>100</td>
</tr>
<tr>
<td></td>
<td>99</td>
</tr>
<tr>
<td></td>
<td>100</td>
</tr>
<tr>
<td>Session 2 images</td>
<td>98</td>
</tr>
<tr>
<td></td>
<td>100</td>
</tr>
<tr>
<td></td>
<td>92</td>
</tr>
<tr>
<td></td>
<td>99</td>
</tr>
<tr>
<td></td>
<td>99</td>
</tr>
<tr>
<td></td>
<td>88</td>
</tr>
<tr>
<td></td>
<td>99</td>
</tr>
<tr>
<td>Overall accuracy</td>
<td>88</td>
</tr>
</tbody>
</table>

<sup>a</sup> ELBDS algorithm depicted here uses test image perturbations of ±5 pixels. Color images are used. It can be noted that with gray images overall accuracy is 95%.

<sup>b</sup> PCA, SIS and SIS<sub>nondiv</sub> results has been adopted from [24]. For LPS results see [28] and for details of SOM-face see [25].

<sup>c</sup> Results not available from the literature.
5. Enhanced LBDS Algorithm

5.4.2 Experiments with other Databases

5.4.2.1 FERET Database

FERET is a large gallery database organised as follows: the gallery set called Fa set is formed of 1,196 face images of unique subjects [103,125,126]. The test set is formed of 1,195 images from probe set called Fb set, 194 images from probe set called Fc set, 722 images from probe called Dup1 set and 234 images from probe called Dup2 set. In this work, the images were cropped to 160 × 160 pixels and then resized to 60 × 60 pixels.

The same set of parameters used in the case of AR database described in Section 5.3.1 are used for simulations. The only difference is in the application of perturbation which in this case is applied to gallery images [see Eq. (5.13)]. In this case, 15 random perturbations in the range of ±5 pixels are applied to either of eye-coordinates and hence account for the localization errors that occur due to improper scaling, rotation and shifts.

Table 5.6 contains the result of the experiment done on FERET database with perturbation and use of preprocessing spatial filters as the differentiating components. It can be seen that localization error compensation done by image perturbation increases the recognition performance considerably, while the use of filters also show a moderate improvement in the recognition accuracy. Further, these results are compared with other algorithms such as SIS and the standard algorithms in the CSU face evaluation system [129]. Weighted LGBPHS algorithm that uses Gabor filters and SIS algorithm that uses Gradient filters have similar as that of the ELBDS algorithm show comparable recognition performance. However, they use much more complex classifiers than ELBDS algorithm. Further, the ELBDS algorithm shows a superior recognition performance for the comparisons done against standard methods shown in Table 5.7.
Table 5.6: Overall top rank recognition performance of ELBDS algorithm with other databases for single image per person problem

<table>
<thead>
<tr>
<th>Databases</th>
<th>Perturbation</th>
<th>No</th>
<th>Yes</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>ELBDS filters</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>No^a</td>
<td>Yes</td>
<td></td>
</tr>
<tr>
<td>AR</td>
<td>88</td>
<td>93</td>
<td>92</td>
</tr>
<tr>
<td>FERET (Fa-Fb)</td>
<td>88</td>
<td>92</td>
<td>95</td>
</tr>
<tr>
<td>FERET (Fa-Fc)</td>
<td>71</td>
<td>78</td>
<td>91</td>
</tr>
<tr>
<td>FERET (Fa-Dup1)</td>
<td>52</td>
<td>57</td>
<td>67</td>
</tr>
<tr>
<td>FERET (Fa-Dup2)</td>
<td>52</td>
<td>57</td>
<td>68</td>
</tr>
<tr>
<td>YALE</td>
<td>94</td>
<td>99</td>
<td>99</td>
</tr>
<tr>
<td>EYALE</td>
<td>80</td>
<td>97</td>
<td>82</td>
</tr>
<tr>
<td>CALTECH</td>
<td>90</td>
<td>94</td>
<td>96</td>
</tr>
</tbody>
</table>

^a This represents the corresponding results on LBDS algorithm.

5.4.2.2 YALE Database

The YALE database contains 165 grayscale images of 15 persons with each person having 11 different images. The conditions they represent are: (1) facial expression: happy, normal, sad, sleepy, surprised, and wink, (2) effect of eye glasses: with glasses, without glasses, and (3) illumination: left-light, right-light, center-light. The images are localized and cropping to a size of 100 × 135 pixels and then resized to 60 × 60 pixels. The first image of each person is taken as the gallery image, while the remaining images form the test images. Again the same set of parameters are used as described in Section 5.3.1 for the simulations. For the ELBDS algorithm, a top rank accuracy of 99% is achieved without applying perturbation and 100% in the case when perturbation is applied as shown in Table 5.6.
Table 5.7: Comparison of top rank recognition performance of ELBDS algorithm with other algorithms on FERET database

<table>
<thead>
<tr>
<th>Methods</th>
<th>Recognition accuracy on FERET database (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Fa-Fb</td>
</tr>
<tr>
<td>ELBDS</td>
<td>98</td>
</tr>
<tr>
<td>Weighted LGBPfHSA</td>
<td>98</td>
</tr>
<tr>
<td>SISb</td>
<td>91</td>
</tr>
<tr>
<td>PCA, MahCosinec</td>
<td>85</td>
</tr>
<tr>
<td>PCA_Euclideanc</td>
<td>74</td>
</tr>
<tr>
<td>Bayesian,MAPc</td>
<td>82</td>
</tr>
<tr>
<td>EBGM, Standardc</td>
<td>88</td>
</tr>
<tr>
<td>LDA_Euclideanc</td>
<td>61</td>
</tr>
</tbody>
</table>

* a Reported by [130].
* b Reported by [24].
* c Simulated using CSU face recognition system [24, 129].

5.4.2.3 Extended Yale Database

The Extended Yale Database (EYALE) database contains serious illumination variation. The database contains 5760 single light source images of 10 subjects each seen under 576 viewing conditions (9 poses \( \times \) 64 illumination conditions). The images with frontal pose is alone selected for the experimental results reported in this chapter. The images are organised into 5 subset based on the illumination conditions illustrated in Fig. 5.3. This classification is made based on the angle of the light-source directions [azimuth and elevation] from the optical axis. Subset one, two, three, four and five are formed of images with illumination angles of: (1) less than 12 degree, (2) between 20 and 25 degrees, (3) between 35 and 50 degrees, (4) between 60 and 77 degrees, and (5) above 77 degrees, respectively. The manual cropping and localization of the images was done to a size of 192 \( \times \) 168 pixels [132]. These images are then resized to 60 \( \times \) 60 pixels to maintain the consistency on the use of ELBDS parameters.

A single face image of each person having an illumination angle of zero degree
Figure 5.3: The azimuth and elevation angles for the 64 illumination conditions in EYALE database. The 5 subsets corresponds to the five different range of illumination conditions formed of images with illumination angles of: (1) less than 12 degree, (2) between 20 and 25 degrees, (3) between 35 and 50 degrees, (4) between 60 and 77 degrees, and (5) above 77 degrees, respectively [133].

is selected to form the gallery set and the remaining 63 images falling in all of the five subsets are selected as test set. The wide variation in illuminations across different subsets and the limitation in training data make this a difficult task. The recognition performance of 97% is achieved [see Table 5.6 and Table 5.8], which is very high recognition performance for such a difficult testing condition.

5.4.2.4 CALTECH Database

The CALTECH database contains 450 face images of 27 or so unique persons with different lighting or expressions or backgrounds. The images are localized
Table 5.8: Recognition performance of ELBDS algorithm for different subsets of EYALE database.

<table>
<thead>
<tr>
<th>Methods</th>
<th>Subset</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1</td>
</tr>
<tr>
<td>ELBDS</td>
<td>100</td>
</tr>
<tr>
<td>SIS [24]</td>
<td>100</td>
</tr>
<tr>
<td>PCA [24]</td>
<td>95</td>
</tr>
</tbody>
</table>

and cropped to a size of $100 \times 135$ pixels and then resized to $60 \times 60$ pixels. The first image of each person in the CALTECH database form the gallery image and the remaining images form the test set. Table 5.6 contains the brief summary of the the results on CALTECH database. It can be seen that the use of filters and localization error compensation improves the recognition accuracy, and is consistent with the results seen in other database.

5.5 Conclusions

The presented ELBDS algorithm utilises preprocessing spatial filters for extracting various spatial details, spatial change detection for feature extraction and local binary decisions for the purpose of classification. ELBDS filters provide increased invariance towards illumination variability. This is evident from the results on EYALE database which contains high degree of illumination changes. This is because, the use of multiple filters enables more information to be extracted as opposed to using a single filter [see Table 5.3]. Further, the high degree of illumination invariance and availability of more texture information provides a robust face recognition algorithm as is shown by the results reported on different databases such as AR, FERET, YALE, EYALE and CALTECH databases. Furthermore, these results outperform or match the best reported or major results, in the category of using single gallery image per person on the used databases.
The ELBDS classifier provides a method for combining the comparisons of the extracted information from the different filters and provides the local binary decisions necessary for its classification. The average similarity measure that is formulated using Eq. (5.8) reinforces the feature matching and reduces the inclusion of false matched features. Applying local binary decisions on average similarity measures further reduces the inclusion of any false matched features. Also, local binary decisions reduces the false matching by treating each variability above a global threshold $\theta$ as a noisy information. For example, the extracted features from the occlusions can be seen as a noisy information. This effect enables the rejection of feature matching between facial and non-facial objects, and is the reason why the presented method shows a superior performance against occluded images [see Table 5.5]. The ability of the ELBDS classifier to reject noisy information helps with increasing the recognition performance against all the natural variabilities. Facial expressions, illuminations and aging are also seen as occlusion of the face and hence is treated as noisy information.

Another aspect, that has been utilised in the design of this algorithm is the idea of system integration based on modularity. Each component in this algorithm can be viewed as a modular unit designed for doing a specific task. And, these basic modules can be replicated over and over again to build more complex systems. For example, local binary decisions, spatial change and preprocessing spatial filters are replicated and applied to compensate for localization errors and to incorporate the use of color images.

In comparison with the other major methods, ELBDS algorithm shows a robust performance against difficult testing conditions. This can be seen through the results on the study of natural variabilities such as illuminations, occlusions, expressions and the variations in time on various databases. The results were achieved using the same set of ELBDS parameters which shows the overall robustness of the algorithm. In summary, ELBDS algorithm provides a simple
and a robust way to extract useful information and classify faces using texture based feature extraction and local binary decisions. High recognition accuracies are achieved using only a single gallery image per person to form the gallery set and can be further improved by using multiple gallery images per person or by inclusion of various learning schemes in real time.
Chapter 6

ELBDS Algorithm in Exampler Based Face Recognition

6.1 Introduction

The successful implementation of face recognition algorithms for single gallery image per person problem was shown in Chapters 2-4. The high recognition performance against such a difficult task enables us to explore further the effect of using multiple gallery images per person to form the gallery set. The increased number of gallery images per person provides more information and hence one can expect higher recognition accuracies than when using a single gallery image per person. However, there are different ways in which the gallery data can be used for comparison: (1) by forming a single feature model (e.g. by taking average) from the set of gallery images of a person [90], and (2) use each gallery image separately for comparison. We will call the first method as average model and the second as exampler model.

In this chapter, recognition performance is analysed for a face recognition method that uses local binary decisions on similarity of features for recognition of faces when there are ordered and unordered multiple number of gallery images.
per person in the gallery set. The average and exampler models are used as training models for comparison of test image with the gallery images. Further, a comparison based on the recognition performance of the designed algorithm against other known best performing algorithms is done. Also, a brief analysis on the effect of using multiple gallery images on localization error compensation and dimensionality reduction is provided. Finally, an application where the method of examplers is used for variability detection is also shown.

6.2 Face Recognition Method

The local binary decisions on similarity method consists of a feature extractor and a local binary decisions based classifier. An Enhanced Local Binary Decision on Similarity (ELBDS) algorithm is formulated for application to multiple training samples per person face recognition problem. In the ELBDS algorithm, the feature extraction stage consists of preprocessing spatial filters followed by spatial change detection and the classifier applies local binary decisions on an average similarity measure. The use of multiple images can be incorporated by either using a single average model for multiple images or a direct exampler model.

6.2.1 Feature Extraction

Any raw intensity image of size $N \times M$ pixels can be denoted as $I(i, j)$, where $(i, j)$ represents the location of pixel. Further, the gallery images is denoted as $I_g^{(k,c)}$, where $k = \{1 \ldots K\}$ is the index for a person, and $c = \{1 \ldots C\}$ is the index of the sample images of the same person. Similarly, a random test image is represented as $I_t$.

Both gallery and test images require to go through the feature extraction stage. As part of feature extraction, preprocessing spatial filters are used to maximize the availability of identity information in a face. The output from
the preprocessing spatial filters are then subjected to spatial intensity change
detection operation using local standard deviation filter. This completes the
feature extraction stage of the proposed algorithm.

Preprocessing spatial filtering is achieved by a linear convolution between a
specified filter window coefficients $w_f$ and raw image $I$. This operation in general
was shown in Eq. 5.1. This is followed by spatial change detection using a local
standard deviation filter across a window of $n \times m$ pixels summarised in Eq. 5.3

Normalization is performed on Eq. (5.3) to form the final feature vector:

$$x^{(p,*)}(i,j) = \sigma^{(p,*)}(i,j)/\bar{\sigma}^{(p,*)}(i,j)$$  \hspace{1cm} (6.1)

where, the local mean $\bar{\sigma}^{(p,*)}(i,j)$ is calculated on a window of features of size
$k \times l$ pixels.

Six preprocessing spatial filters are used in the proposed method. These
filters and its corresponding outputs are shown in Fig. 6.1 (a)-(f), while its
corresponding normalized spatial change is shown in Fig. 6.1 (g)-(l).

### 6.2.2 Classification

The classification process involves the comparison between the feature vectors of
a test image with that of gallery images. One of the simplest way to form such
comparison is by taking normalized absolute difference between the features (as
shown in Chapters 3, 4, 5), called as similarity measure vector. This feature
comparison operation is summarized as:

$$\delta^{(p,k,c)}(i,j) = \frac{|x^{(p,k,c)}_g(i,j) - x^{(p)}_t(i,j)|}{\min(x^{(p,k,c)}_g(i,j), x^{(p)}_t(i,j))}$$  \hspace{1cm} (6.2)

where, $x^{(p,k,c)}_g$ represents the normalized feature vector formed by applying
$p^{th}$ spatial filter on $c^{th}$ sample image of $k^{th}$ person in the gallery. A comparison
of a test feature vector with a gallery image having a index $k$ will result in $p$
similarity measure vectors due to the $p$ preprocessing spatial features employed
Figure 6.1: Illustration of preprocessing spatial filters and spatial change detection applied to a face image in AR database. The weights of the preprocessing spatial filters and its corresponding outputs are shown in images labeled (a) to (f). The spatial change detection is applied on each of these outputs and is shown in images labeled (g) to (l).

during the feature extraction process. To simplify the calculation and to reinforce the identity information the $p$ similarity measure vectors are averaged to form a single average similarity measure vector given by:

$$\hat{\delta}(k,c) = \frac{1}{P} \sum_{p=1}^{P} \delta(p,k,c)(i,j)$$  \hspace{1cm} (6.3)

Application of local binary decisions on $\hat{\delta}(k,c)$ using a global threshold $\theta$ results in a binary decisions vector $B^{(k,c)}$ represented as:

$$B^{(k,c)}(i,j) = \begin{cases} 
1 & \hat{\delta}(i,j) < \theta \\
0 & \hat{\delta}(i,j) \geq \theta 
\end{cases}$$  \hspace{1cm} (6.4)

This binary decisions vector is used form a global similarity score $S^{(k,c)}$ given by:

$$S^{(k,c)} = \sum_{i=1}^{N} \sum_{j=1}^{M} B^{(k,c)}(i,j)$$  \hspace{1cm} (6.5)
The comparison of a test image with \( d \times k \) gallery images results in \( d \times k \) similarity scores and the maximum value among all such scores will represent the best match of the test image in the gallery:

\[
kc^* = \arg \max_{k,c} S_g^{(k,c)}
\]  

(6.6)

### 6.2.3 Compensation of Localization Errors

Another, important practical aspect in face recognition is the localization error that occur due to any improper detection and alignments. In most practical cases, a perfect alignment is difficult to achieve, so it is logical to add a scheme to compensate for such errors. In practice, localization of faces is done with respect to eye coordinates of the face image in majority of face detection schemes. This knowledge is used to form a simple compensation method by perturbating location of eye coordinates to generate scaled, shifted and rotated versions of an image. These perturbations can be applied to the gallery or the test image to compensate for variations in shifts, scale and rotation. Any \( u \) number of perturbations on an image will result in \( u \) similarity measure, and the one with the maximum value is chosen as the representative score for that comparison. In this way small variations in localization errors that can reduce the recognition performance is compensated.

### 6.3 Experimental Results

#### 6.3.1 Equal Number of Training Samples Per Person

One particular situation in multiple training samples per person problem occurs when there are equal number of training samples per person with same conditions of natural variability for all the persons in the gallery. This method of setting up
the gallery is perhaps the easiest way to achieve high recognition performance
due to the equal probability of match on the identity information of a test image
with that in the gallery images. Further, within this scheme, the gallery samples
can be setup in two ways: (1) use all the training samples per person directly
for comparison (exampler method), and (2) form a single model gallery image
for a person from all the training samples for a person (average method).

For the experimental analysis, on all the databases following are the values
of parameters used in the algorithm for the value of image/feature vector size
kept at a range from $10 \times 10$ pixels to $90 \times 90$ pixels: (1) standard deviation
filter size is $3 \times 3$ pixels, (2) local mean normalization window size for forming
normalized features is kept at $30 \times 30$ pixels, (4) the value of global threshold
is 0.25 and perturbations of $\pm 5$ pixels are applied in horizontal, vertical and
diagonal directions.

AR database [1, 2] with images of 100 persons is used for the simulations
reported in this chapter. For each person there are 26 different images representing 13
conditions over two sessions. Seven images of each person are selected
as training samples to form the gallery (see Fig. 6.2) and the remaining 19 images
of each person is used for testing the recognition performance of the algorithm.
Figure 6.3: Graphical illustration showing the recognition performance of the proposed method with variation in dimensionality of feature vectors. This shows the specific situation when the number of training samples used for creating the gallery is same for all the persons. AR database is employed for these simulation. Further this also shows a comparison of single model approach such as average feature model against multiple feature models such as exampler features.
6. ELBDS Algorithm in Exemplar Based Face Recognition

Figure 6.4: Graphical illustration showing the time it takes for the classifier for a comparison of a test with all images in the gallery at different feature vector dimensions.
From Fig. 6.3 it can be seen that exampler method performs better than average method for all the shown variation in the dimensionality. However, exampler method require $c$ times more amount of memory and comparisons as opposed to average method and hence results in larger computational time as shown in Fig. 6.4. Further, it can be seen from Fig. 6.3 that exampler method shows robust recognition performance than average method with respect to variations in dimensionality of feature vector.

Table 6.1 shows the recognition performance and overall robustness of proposed method using color images and perturbations. For AR [1, 2], ORL, FERET [103,125], EYALE [132,133], YALE [78] and CALTECH databases the reported results are for the feature size of $60 \times 60$ pixels, $40 \times 40$ pixels, $60 \times 60$ pixels, $80 \times 80$ pixels, $60 \times 60$ pixels and $80 \times 80$ pixels respectively. ORL database which contains pose variation of 40 persons is also tested to benchmark the performance, and clearly show high recognition performance. FERET database with images of 200 persons are used and each person has 3 photos representing different natural variability. EYALE database has 64 photos with different illumination variation on each of the 10 persons in the database. This database is often used to benchmark the face recognition performance against serious illumination variations. YALE database has photos of 15 persons under 11 different natural variability. CALTECH face database has photos of 28 persons under random background and in natural lighting conditions. Clearly, from Table 6.1 the proposed method show high recognition performance in all the tested databases, which confirms the overall robustness of the method.

### 6.3.2 Unequal Number of Training Samples Per Person

Yet another problem in face recognition with multiple training samples per person is formed when number of training samples per person is different. However, this problem is closely related to how learning occurs in primates where the
Table 6.1: Recognition performance across different databases on the proposed method when the number of training samples in the gallery is fixed for every person.

<table>
<thead>
<tr>
<th>Database</th>
<th>Samples per person</th>
<th>Recognition Accuracy(%)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Gallery</td>
<td>Test</td>
</tr>
<tr>
<td>AR</td>
<td>7</td>
<td>19</td>
</tr>
<tr>
<td>ORL</td>
<td>5</td>
<td>5</td>
</tr>
<tr>
<td>FERET</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>EYALE</td>
<td>7</td>
<td>57</td>
</tr>
<tr>
<td>YALE</td>
<td>5</td>
<td>6</td>
</tr>
<tr>
<td>CALTECH</td>
<td>5</td>
<td>1-15</td>
</tr>
</tbody>
</table>

availability of information for training a face is different from another. Furthermore, for majority of biometric application such scenarios are mostly expected to occur.

Figure 6.5 shows the effect of monotonic increase in training samples in creating a gallery set for all the persons. It can be clearly seen that an increase in available information during training increases the recognition accuracy considerably. Further, this also means that the increased use of memory for storing any distinct identity information results in better recognition performance and stability. This is substantiated from the simulation results shown in Fig 6.6. These results show that using exampler method, involving multiple representation of a face provides greater stability and recognition performance as opposed to a single model approach.

6.3.3 Comparison with Other Methods

The proposed method using examplers is compared with some of the other best performing and well known algorithms reported in the area of face recognition. This comparison is organised based on the databases used to report the result.
Figure 6.5: A graphical illustration showing the effect of using multiple training samples. The simulation is done by exemplar feature comparisons using gray scale images of size $60 \times 60$ pixels in the AR database and the proposed method is used without compensating for localization errors.
Figure 6.6: Graphical illustration showing the recognition performance of the proposed method with variation in dimensionality of feature vectors when the number of training samples used for creating the gallery are randomly selected and are not same for all the persons. Using AR database this simulation also shows a comparison of average feature model against exampler feature model.
Table 6.2: The comparison of recognition performance of the proposed ELBDS method for multiple training samples person face recognition problem with other algorithms using AR database.

<table>
<thead>
<tr>
<th>Test images</th>
<th>ELBDS</th>
<th>SIS</th>
<th>LDA</th>
<th>DCV</th>
<th>PCA</th>
</tr>
</thead>
<tbody>
<tr>
<td>100</td>
<td>99</td>
<td>65</td>
<td>73</td>
<td>58</td>
<td></td>
</tr>
<tr>
<td>100</td>
<td>100</td>
<td>51</td>
<td>60</td>
<td>56</td>
<td></td>
</tr>
<tr>
<td>100</td>
<td>100</td>
<td>57</td>
<td>47</td>
<td>14</td>
<td></td>
</tr>
<tr>
<td>100</td>
<td>99</td>
<td>50</td>
<td>47</td>
<td>10</td>
<td></td>
</tr>
<tr>
<td>100</td>
<td>100</td>
<td>91</td>
<td>88</td>
<td>84</td>
<td></td>
</tr>
<tr>
<td>100</td>
<td>99</td>
<td>82</td>
<td>82</td>
<td>78</td>
<td></td>
</tr>
<tr>
<td>99</td>
<td>99</td>
<td>86</td>
<td>84</td>
<td>77</td>
<td></td>
</tr>
<tr>
<td>100</td>
<td>97</td>
<td>37</td>
<td>45</td>
<td>27</td>
<td></td>
</tr>
<tr>
<td>99</td>
<td>97</td>
<td>22</td>
<td>33</td>
<td>27</td>
<td></td>
</tr>
<tr>
<td>96</td>
<td>92</td>
<td>25</td>
<td>24</td>
<td>6</td>
<td></td>
</tr>
<tr>
<td>96</td>
<td>95</td>
<td>24</td>
<td>18</td>
<td>5</td>
<td></td>
</tr>
<tr>
<td>Overall</td>
<td>99</td>
<td>98.1</td>
<td>56.5</td>
<td>57.4</td>
<td>44.7</td>
</tr>
</tbody>
</table>
6. ELBDS Algorithm in Exampler Based Face Recognition

6.3.3.1 AR Database

The following methods are used for comparison against the proposed ELBDS algorithm: SIS [24], LDA [78], DCV [134] and PCA [10]. Clearly, from Table 6.2 it can be seen that ELBDS algorithm outperforms all other methods. In this simulation 7 images per person is used for forming the gallery and 19 images per person are used for testing. It can be also noted that AR database contains occlusions of faces which is often considered as a difficult task for recognition. Also this simulation shows the robustness of ELBDS algorithm against different natural variability under the same pose.

6.3.3.2 ORL, FERET and YALE Databases

As seen from Table 6.3, MGFR-2D^2PCA [135] and 2D^2PCA [136] are used for comparison against presented method. First five images of each person in ORL database is used as training samples, while the remaining five are used for testing. In the case of the FERET database, any 2 images are selected as the training sample, while the remaining image forms the test sample. For each person in the Yale database, the first 5 images are used to form the training sample and the remaining 6 images forms the test samples. It can be seen from the reported results that the presented method has very high recognition accuracy and is robust against different databases. Further, it is worth while to mention that in the case of YALE database 100 percent recognition accuracy is achieved even with single training sample per person. In terms of natural variability these databases include variations in pose, expressions and illumination.

6.3.3.3 EYALE Database

Results on EYALE database (see Table 6.4) shows the recognition performance of the presented method against other best performing methods such as 9PL_{real} [132], Cone cast [133] and SIS [24]. In each of these methods, the images in
Table 6.3: The comparison of recognition performance of the proposed ELBDS method for multiple training samples person face recognition problem with other algorithms.

<table>
<thead>
<tr>
<th>Database</th>
<th>Top rank recognition accuracy (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>ELBDS</td>
</tr>
<tr>
<td>ORL</td>
<td>99.5</td>
</tr>
<tr>
<td>FERET</td>
<td>99.5</td>
</tr>
<tr>
<td>YALE</td>
<td>100.0</td>
</tr>
</tbody>
</table>

Table 6.4: The comparison of recognition performance of the proposed ELBDS method for multiple training samples person face recognition problem with other algorithms in EYALE database.

<table>
<thead>
<tr>
<th>Method</th>
<th>Recognition accuracy (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Subset</td>
</tr>
<tr>
<td></td>
<td>2</td>
</tr>
<tr>
<td>ELBDS</td>
<td>100</td>
</tr>
<tr>
<td>SIS</td>
<td>100</td>
</tr>
<tr>
<td>Cones-cast</td>
<td>100</td>
</tr>
<tr>
<td>9PL_\text{real}</td>
<td>100</td>
</tr>
<tr>
<td>LDA</td>
<td>100</td>
</tr>
<tr>
<td>DCV</td>
<td>100</td>
</tr>
<tr>
<td>PCA</td>
<td>90</td>
</tr>
</tbody>
</table>

subset 1 consisting of 7 images with a zero degree angle of the light-source from the optical axis are used as the training samples. Further, it can be noted that the shown results on cone cast and 9PL\_\text{real} are obtained by using first four subsets, while for SIS and ELBDS all the five subsets are used. Clearly, from Table 6.4 the comparison with well known method like cone cast and 9PL\_\text{real}, the presented method shows very similar recognition performance. This also shows the near invariant recognition performance of the proposed method against serious illumination changes.
6.3.4 Application of Examplers in Variability Detection

Examplers also provide knowledge on condition or natural variability. This prior knowledge of the gallery of examplers helps with detections of facial expressions, variations in time, aging and pose. In order to demonstrate this, 13 images taken on session 1 in the AR database is used as the example gallery images, while the remaining 13 images from session 2 are used as test. The 13 images are grouped based on the knowledge of the conditions as: (1) Neutral, (2) Expression, (3) Illumination, (4) Eye occlusion and (5) Mouth occlusion. The test image when compared with the images in the gallery results in a set of similarity scores. These scores are ranked and the associated natural variability is tagged to the test image.

Table 6.5 shows the recognition performance of the proposed face recognition algorithm in the detection of natural variability in the image. The detection of occlusion at rank one seems to be most difficult task. This is due to the fact that mouth occlusion images contain the least amount of identity information and will result in lower values of similarity score. This is a relative disadvantage of using a global similarity score, however a possible way to improve the recognition performance is by using local approaches such as the calculation of region-wise similarity scores and weights. The ability of the proposed method to detect natural variability and recognise faces with high accuracies makes it useful in the various applications of automatic data tagging of face images.

6.4 Conclusions

In summary, an example based local binary decisions on similarity algorithm is presented that is successfully applied to multiple training samples per person face recognition problem. The relative advantage of using examplers in comparison with single model approach such as averaging is shown. In the case
Table 6.5: Recognition accuracy of the proposed ELBDS algorithm in the detection of natural variability in the face images.

<table>
<thead>
<tr>
<th>Rank</th>
<th>Neutral</th>
<th>Expression</th>
<th>Illumination</th>
<th>Eye occlusion</th>
<th>Mouth occlusion</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>83.0</td>
<td>87.4</td>
<td>96.0</td>
<td>99.0</td>
<td>73.0</td>
</tr>
<tr>
<td>2</td>
<td>98.0</td>
<td>100.0</td>
<td>99.0</td>
<td>100.0</td>
<td>89.7</td>
</tr>
<tr>
<td>3</td>
<td>99.0</td>
<td>100.0</td>
<td>99.7</td>
<td>100.0</td>
<td>96.7</td>
</tr>
<tr>
<td>4</td>
<td>100.0</td>
<td>100.0</td>
<td>100.0</td>
<td>100.0</td>
<td>100.0</td>
</tr>
</tbody>
</table>

of exampler based approach, the increased usage of memory enables the use of more identity information as opposed to single model approach. Although single model approach is computationally less expensive, the use of examplers enables a stable performance even with reduced feature dimensionality. The presented method outperforms other major algorithms in overall robustness across various natural variabilities. This is attributed to the use of preprocessing spatial change features and the use of local binary decisions classifier. Further, the use of multiple training samples helps in compensation of natural variability and increases the probability for a true match. In addition, a useful aspect of this method is its ability to detect natural variability which sets this method apart from its counterparts. Finally, from the results it is evident that any increase in the number of examplers will make the recognition performance higher and more stable across variations in natural variability and feature dimensionality.
Chapter 7

Conclusions

7.1 Summary of Contributions

This thesis demonstrates that complex recognition decisions can be made by aggregating simple local decisions. As shown through the various face recognition algorithms using the concepts of local binary decisions and spatial intensity change, the proposed method is general, robust and simple. Since, the method only involves operations like convolution and summation, makes it easy to implement as a hardware or software algorithm. High recognition performance on a large number of face image databases demonstrate the effectiveness and robustness of the approach even to difficult face recognition task involving single gallery image per person.

7.1.1 List of Contribution

The brief summary of chapter wise contribution to knowledge and insights provided in this thesis are listed as follows:

1. Local binary decisions based face recognition (Chapter3):

   (a) The principle of firing of neurons is investigated to develop the concept
of local binary decisions.

(b) The *spatial intensity changes* in face image is found to be the main visual cue required for recognition.

(c) The concept of local binary decisions and spatial intensity changes are presented and successfully implemented by forming a baseline LBDS algorithm for face recognition.

(d) The concept of local binary decisions is verified by using different implementation of similarity measures, while that of spatial intensity changes is verified using different implementation of spatial filters.

(e) The overall robustness of the LBDS algorithm is verified by using various databases having different natural variability and with limited training data.

2. Analysis of LBDS Algorithm (Chapter 4):

(a) The main contributing factors for high recognition accuracy are studied. Local binary decisions and spatial change features are found to be the major contributing factors that enable high recognition performance.

(b) It is found that localization error compensation, color, resolution and normalization of features and similarity measures helps to increase the recognition performance and robustness.

(c) The LBDS algorithm with optimized parameters is compared with other major or best performing algorithms. The proposed method shows good recognition performance in comparison with these methods.

3. Enhanced LBDS algorithm (Chapter 5):
7. Conclusions

(a) The importance of extracting the information from the image is shown through the use of various preprocessing spatial filters.

(b) The use of different preprocessing spatial filters provides details of the faces that is suppressed by natural variability. It is shown that preprocessing is required to maximize the detection of all the significant spatial change features useful to ascertain the identity of the face.

(c) It is shown that local binary decisions on average similarity measure provides increased robustness to noisy information and hence natural variability.

(d) Through the design and implementation of the ELBDS algorithm, a successful implementation of a face recognition algorithm based on the principle of modularity is presented.

(e) The presented method shows superior recognition performance in comparison with other best performing or well known algorithms applied to single gallery image per person problem on various publicly available databases.

4. ELBDS algorithm in exampler based face recognition (Chapter 6):

(a) The principle of modularity and the idea of extracting the spatial information from the image is used to develop an exampler based face recognition algorithm.

(b) Through the simple example of using multi-model approach such as examplers and single-model approach such as image average, the relative advantage of using multi-model representation of face is shown.

(c) The ability of exampler based face recognition to provide increased robustness and stability against reduction in feature dimensions is demonstrated.
(d) The example based face recognition algorithm shows that using very simple concepts of local binary decisions, spatial change features and principle of modularity a complex system is build in a hierarchical manner.

7.2 Future Directions

Here, a perspective on the future directions of the proposed approach at a higher level is suggested that aims at pattern recognition research in general.

⇒ Texture analysis for image enhancement

The proposed techniques showed the importance of extraction of texture based features. Clearly, increased number of features helps in the increased robustness to natural variability. The texture analysis can play a key role in real-time processing like in scanners, and document processing instruments. The application of spatial change detections can be further utilized in the document processing stages in identifying characters irrespective of languages used and as a means of normalization when multiple preprocessing spatial filters are used. Further investigation on developing an affine invariant feature representation using the spatial features can be done for use in segmentation and detection.

⇒ Classification using local binary decisions

It can be seen that in this thesis, the local binary decisions are made using single threshold. However, multiple threshold scenarios can be investigated for use with any special applications of vision science. The simple way of achieving classification using a threshold helps to form a very hardware friendly algorithm. A general local binary decisions classification scheme using VLSI logic circuits can be formed and implemented for this cause. Further, local binary decisions can be investigated in the formation of knowledge based system and can be used
to provide a simple means to implement an early model of intelligence. Various other studies on cognitive activity and sensory interactions based on local binary decisions can be investigated.

⇒ *Emulating the neural sensory processing*

Through the experiments discussed in this thesis, it is clear that simple operations like decision thresholding and signal variability detection when emulated from the biology, can actually produce some outstanding results. Using the same approach, various other sensory processing and useful recognition mechanisms and be emulated in the future.
Bibliography


[81] A. M. Martinez, “Recognizing imprecisely localized, partially occluded and
eexpression variant faces from a single sample per class,” IEEE Trans. on
Pattern Analysis and Machine Intelligence, vol. 24, no. 6, 2002.

jects from appearance,” International Journal of Computer Vision, vol. 14,

recognition,” in Proceedings of International Workshop on Automatic Face
and Gesture Recognition, Zurich, Switzerland, pp. 80–85, 1995.

[84] D. Beymer and T. Poggio, “Face recognition from a single example view,”
in Proceedings of 5th International Conference on Computer Vision, Bo-

[85] V. Blanz and T. Vetter, “Face recognition based on fitting a 3D morphable
model,” IEEE Transactions on Pattern Analysis and Machine Intelligence,

reconstruction for face recognition,” Pattern Recognition, vol. 38, no. 6,

recognition as people age,” in IEEE International Conference on Computer

[88] N. Ramanathan and R. Chellappa, “Face verification across age progress-
sion,” in IEEE Conf. Computer Vision and Pattern Recognition (CVPR),


