Automatic approaches towards vehicle make and model recognition

This item was submitted to Loughborough University’s Institutional Repository by the/an author.

Additional Information:


Metadata Record: https://dspace.lboro.ac.uk/2134/33897

Publisher: © Iffat Zafar

Rights: This work is made available according to the conditions of the Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International (CC BY-NC-ND 4.0) licence. Full details of this licence are available at: https://creativecommons.org/licenses/by-nc-nd/4.0/

Please cite the published version.
Please note that fines are charged on ALL overdue items.
AUTOMATIC APPROACHES TOWARDS VEHICLE MAKE AND MODEL RECOGNITION

by

Iffat Zafar

A doctoral thesis
submitted in partial fulfilment of the requirements
for the award of

Doctor of Philosophy

Department of Computer Science
Loughborough University
September 2008

© by Iffat Zafar 2008

Supervisors: Dr. Eran Edirisinghe
Dr. Serpil Acar

Director of Research: Dr. Chris Hinde
ABSTRACT

Vehicle Make & Model Recognition (MMR) systems provide means for improving the level of identity provided by Automatic Number Plate Recognition (ANPR) systems. Although a number of vehicle MMR systems have been proposed in literature the design of a robust algorithm still remains an open research problem. The research presented in this thesis aims to resolve some of the outstanding challenges in vehicle MMR such as rotation, translation, scale and illumination variance, occlusion, computational complexity and functional limitations.

Four novel algorithms for vehicle MMR are proposed. The first uses Two Dimensional Linear Discriminant Analysis (2DLDA) as an appearance based approach to vehicle MMR. This approach provides improved robustness against illumination variations and occlusion as compared to state-of-the-art Principle Component Analysis (PCA) based vehicle MMR systems. Further the computational cost is minimised by the selected use of Eigenvalues. The second approach uses Scale Invariant Feature Transforms (SIFT) in vehicle MMR. It uses a novel interest point (SIFT keypoint) matching strategy that is capable of significantly improving the recognition accuracy as compared to the use of the traditional approach to SIFT keypoint matching. This approach provides robustness against scale, translation and limited rotation. It is later extended by using Adaptive Boosting (AdaBoosting) for feature selection. The use of AdaBoosting enables the identification of SIFT features which are most representative of any particular vehicle make and model, thus reducing computational cost of SIFT keypoint matching and increasing the recognition accuracy due to noise reduction and effective background removal. The fourth approach uses local texture features in the lowpass sub-band and the oriented image edge features across the scales of directional resolutions of Contourlet transformed image as features, rather than the direct use of Contourlet transform coefficients as features, in vehicle MMR. The use
of above localised directional feature maps significantly improves the recognition accuracy when compared to using the standard deviations of the Contourlet sub-bands in image matching.

The proposed algorithms have been tested on a database of 200 images consisting of eight distinct car make-models, providing recognition accuracies ranging from 90-94%. The thesis further highlights possible extensions and improvements to the proposed approaches for vehicle MMR.

Iffat Zafar, September 2008
To my beloved Parents

(Ulfat Begum & Raja Zafar Ali Khan)
ACKNOWLEDGEMENT

This thesis arose in part out of three years of research. By that time, I have worked with a great number of people whose contribution in assorted ways to the research and the making of the thesis deserved special mention. It is a pleasure to convey my gratitude to them all in my humble acknowledgment.

The chain of my gratitude begins with name of Almighty Allah (s.w.t), the Most Gracious, and the Most Merciful whose blessings are always with me and due to which I got the courage and knowledge to accomplish my tasks.

I shall like to express my deep and sincere gratitude to my supervisors, Dr Eran Edirisinghe and Dr Serpil Acar for their supervision, invaluable assistance, advice, abundant help and guidance from the very early stage of this research as well as giving me extraordinary experiences throughout the work. Their wide knowledge and logical way of thinking have been of great value for me. Their detailed and constructive comments, technical and editorial advice was essential to the completion of this thesis. Above all and the most needed, they provided me unflinching encouragement and support in various ways. I am indebted to them more than they know.

I thank my director of research, Dr.Chris Hinde for his time, encouragement and suggestions. I am deeply grateful to Dr Helmut Bez for imparting the essential mathematical knowledge and help in understanding various mathematical concepts whenever required. I am also thankful to the technical and clerical staff members of Department of Computer Science, Loughborough University.

I shall also like to convey thanks to the Innovative Manufacturing and Construction Research Centers (IMCRC) Loughborough University, and Engineering and
Physical Sciences Research Council (EPSRC), UK, for providing funding to complete my PhD.

Further I wish to acknowledge the support of Dr V. Petrovic, University of Manchester, for providing the car image database used for the experiments in this thesis, Ms. Alex Weekes from Thatcham, UK, for providing data for extending the car image database, Andrea Vedaldi, UCLA Vision Lab, Department of Computer Science, University of California, for providing the source code for SIFT, Saquib Sarfraz for sharing his knowledge, ideas and valuable discussions about my PhD research and Usman Zakir for sharing ideas and proof reading my thesis.

Thanks to my colleagues at Digital Imaging Lab (DIL) and Biomechanics and Injury Prevention (BmIP) research group, who shared their knowledge, provided useful feedback on my work, provided a nice friendly environment to work and made my three years enjoyable and memorable. Special thanks to my friends for their prayers, care, support, motivation, and time throughout my PhD.

Where would I be without my family? My parents deserve special mention for their unconditional, inseparable support, prayers and endless love. My Father, Raja Zafar Ali Khan, in the first place is the person who put the fundament in my learning character, showing me the joy of intellectual pursuit ever since I was a child. My Mother, Ulfat Begum, is the one who sincerely raised me with her caring and gentle love and prayed day and night for my success. My special gratitude to my brothers: Majid, Imran and Kamran, my sisters: Farhat, Sarwat and Shabnum, my sister-in-law: Shafaq, my brothers-in-law: Kaleem and Ali, my nieces and my nephews. Thanks for being supportive, loving and caring.

Iffat Zafar

26th September 2008
TABLE OF CONTENTS

ABSTRACT.................................................................III
DEDICATION...............................................................V
ACKNOWLEDGMENT.....................................................VI
TABLE OF CONTENTS....................................................VIII
LIST OF FIGURES AND TABLES.................................XV
ABBREVIATIONS AND NOTATIONS..........................XX

CHAPTER 1:
AN OVERVIEW.........................................................1-8

1.0 INTRODUCTION..................................................1
1.1 RESEARCH MOTIVATION.......................................2
1.2 AIM AND OBJECTIVES.........................................4
1.3 CONTRIBUTIONS OF RESEARCH............................5
1.3.1 Use of Two Dimensional Statistical Linear Discriminant Analysis (2DLDA) in Vehicle MMR...........5
1.3.2 Use of Scale Invariant Feature Transforms (SIFT) in Vehicle MMR........................................6
1.3.3 Use of Adaptive Boosting in Feature Selection for Vehicle MMR.............................................6
1.3.4 Use of Contourlet Domain Localised Directional Feature Maps in Vehicle MMR......................7
1.4 ORGANISATION OF THESIS.................................7
# TABLE OF CONTENTS

## CHAPTER 2:

**LITERATURE REVIEW: VEHICLE MAKE AND MODEL RECOGNITION (MMR) TECHNIQUES** ................................................................. 9-27

- **2.0 INTRODUCTION** ................................................................. 9
- **2.1 DETECTION OF CARS IN SCENES** ........................................... 10
- **2.2 VEHICLE MMR ALGORITHMS** ............................................... 12
  - **2.2.1 Region of Interest (ROI) Detection** ..................................... 13
  - **2.2.2 Feature Detection, Extraction and Classification** .................... 15
- **2.3 FACE RECOGNITION: LITERATURE** ...................................... 22
- **2.4 SUMMARY AND CONCLUSION** .............................................. 26

## CHAPTER 3:

**RESEARCH BACKGROUND** .......................................................... 28-79

- **3.0 INTRODUCTION** ................................................................. 28
- **3.1 STATISTICAL PREMILINARIES** ........................................... 28
- **3.2 PRINCIPAL COMPONENT ANALYSIS (PCA)** ............................ 30
  - **3.2.1 PCA for Two Variables** .................................................. 30
  - **3.2.2 Computing Principal components of PCA** .......................... 33
  - **3.2.3 The Use of PCA in Feature Extraction and Classification** ........ 34
- **3.3 LINEAR DISCRIMINANT ANALYSIS (LDA)** .............................. 37
  - **3.3.1 Theory of LDA** ............................................................ 39
### TABLE OF CONTENTS

3.4 **TWO DIMENSIONAL LINEAR DISCRIMINANT ANALYSIS (2DLDA)** .................................................. 41

3.4.1 2DLDA Algorithm for Feature Extraction .......................................................... 41

3.4.2 2DLDA and Singularity .................................................................................. 44

3.5 **SCALE IN Variant FEATURE TRANSFORM (SIFT)** ............................................ 45

3.5.1 Detection of Scale-space Extrema ........................................................................ 45

3.5.2 Keypoint Localisation ....................................................................................... 48

3.5.3 Orientation Assignment ..................................................................................... 48

3.5.4 Keypoint Descriptor Computation ........................................................................ 49

3.5.5 Invariance Properties of SIFT ............................................................................. 50

3.6 **CURVELET TRANSFORMS** ............................................................................. 50

3.6.1 Stages of Curvelet Transforms ........................................................................ 51

3.6.2 Why Use Curvelets? ......................................................................................... 55

3.6.3 Challenges in Using Curvelets .......................................................................... 56

3.7 **CONTOURLET TRANSFORMS** ........................................................................ 56

3.7.1 Discrete Domain Construction .......................................................................... 57

3.7.2 Why Contourlets? .............................................................................................. 60

3.8 **SUPPORT VECTOR MACHINES (SVM)** .......................................................... 61

3.8.1 Basic Theory of SVM ....................................................................................... 61

3.8.2 Advantages of SVM ......................................................................................... 68

3.9 **CROSS VALIDATION** ....................................................................................... 68

---

**Automatic Approaches Towards Vehicle Make and Model Recognition** X
TABLE OF CONTENTS

3.9.1  $n$-fold Cross Validation ..................................................69

3.10  $k$-NEAREST NEIGHBOUR CLASSIFIER ($k$-NN) ..........................69

3.10.1 $k$-NN Algorithm ...............................................................70

3.10.2 Strengths and Weaknesses of $k$-NN Classification ..................71

3.11 BOOSTING .................................................................71

3.11.1 Adaptive Boosting (AdaBoost) ...........................................72

3.11.2 AdaBoost Algorithm ..........................................................72

3.11.3 Advantages of AdaBoost ....................................................74

3.12 DATASETS .................................................................74

3.13 SUMMARY .................................................................78

CHAPTER 4:

USE OF TWO DIMENSIONAL STATISTICAL LINEAR DISCRIMINANT
ANALYSIS (2DLDA) IN VEHICLE MMR ........................................80-93

4.0 INTRODUCTION .............................................................80

4.1 MOTIVATION ...............................................................81

4.2 PROPOSED METHOD ..........................................................82

4.2.1 Vehicle Front Extraction and Normalization ..........................84

4.2.2 Processing the Training Image Set ....................................84

4.2.3 Processing the Test Images .............................................85

4.2.4 Classification ...............................................................85
## TABLE OF CONTENTS

4.3 EXPERIMENTAL RESULTS AND ANALYSIS ........................................ 86

4.4 SUMMARY AND CONCLUSION ......................................................... 92

CHAPTER 5:

**USE OF SCALE INVARIANT FEATURE TRANSFORM (SIFT) IN VEHICLE MMR** .......................................................... 94-101

5.0 INTRODUCTION ........................................................................ 94

5.1 MOTIVATION ............................................................................ 94

5.2 PROPOSED APPROACH .......................................................... 96

5.2.1 Pre-processing the Image Database ....................................... 97

5.2.2 Processing the Test Images ................................................... 98

5.2.3 Matching ............................................................................ 98

5.2.4 Classification ...................................................................... 99

5.3 EXPERIMENTAL RESULTS & ANALYSIS ................................... 99

5.4 CONCLUSION AND FURTHER WORK ....................................... 101

CHAPTER 6:

**USE OF ADAPTIVE BOOSTING IN FEATURE SELECTION FOR VEHICLE MMR** .................................................. 102-117

6.0 INTRODUCTION ......................................................................... 102

6.1 RESEARCH MOTIVATION ........................................................ 103

6.2 PROPOSED APPROACH .......................................................... 104

Automatic Approaches Towards Vehicle Make and Model Recognition  

XII
## TABLE OF CONTENTS

6.2.1 Dataset Preparation ................................................................. 105

6.2.2 Feature Detection ................................................................. 106

6.2.3 Feature Selection ................................................................. 107

6.2.4 Interest Point Matching ......................................................... 109

6.3 EXPERIMENTAL RESULTS & ANALYSIS ..................................... 109

6.4 CONCLUSION ................................................................. 117

CHAPTER 7:

USE OF CONTOURLET DOMAIN LOCALISED DIRECTIONAL FEATURE MAPS IN VEHICLE MMR ........................................... 118-134

7.0 INTRODUCTION ................................................................. 118

7.1 RESEARCH MOTIVATION ......................................................... 118

7.2. PROPOSED APPROACH .......................................................... 121

7.2.1 Pre-Processing of Test and Training Images .............................. 122

7.2.2 Contourlet Decomposition ..................................................... 123

7.2.3 Features Extraction .......................................................... 124

7.2.4 Dimensionality Reduction (Optional) ...................................... 125

7.2.5 Support Vector Machine Based Classification .......................... 125

7.3. EXPERIMENTAL RESULTS & ANALYSIS .................................. 126

7.4. CONCLUSION ................................................................. 134
TABLE OF CONTENTS

CHAPTER 8:

CONCLUSIONS.................................................................135-140

8.0 INTRODUCTION.............................................................135

8.1 CONCLUSION OF THE THESIS..........................................136

8.2 FUTURE WORK.............................................................138

REFERENCES.................................................................141-154

APPENDIX.............................................................................155-156

A  SCHOLARLY CONTRIBUTION...........................................155
LIST OF FIGURES AND TABLES

FIGURES

Figure 2.1  Car Geometry measured in space of number-plate, (a) [23], (b) [27] ...................................................... 14

Figure 3.1  Axis of maximum scatter ................................................. 31

Figure 3.2  Scatter or variance along x-axis ............................... 31

Figure 3.3  'Natural' or 'Principal' axis system, relative to X, Y system for the dataset. The origin of Principal system is at \((x, y)\) in the original system .................................................. 32

Figure 3.4  Data set represented in its Principal axis system ...... 32

Figure 3.5  Ellipse in arbitrary and principal axis system .......... 33

Figure 3.6  Visualization: PCA vs. LDA for a two class problem where data for each class lies near a linear subspace [52] .......... 39

Figure 3.7  Gaussian blurred images at different scales and their respective DOG images [58] .................................................. 46

Figure 3.8  Local extrema detection, the pixel 'x' at \(s_{\pi}\) is compared to total of 26 neighbouring pixels at same level and levels above and below [58] .................................................. 47

Figure 3.9  Left: Keypoint (represented by a black dot) surrounded by neighbouring points. Arrows show the orientations of points. Right: Orientation histogram of magnitude sums for all the points within a window around the keypoints [118] ........... 48

Figure 3.10  Left: the gradient magnitude and orientation at a sample point in a square region around the keypoint location. These are weighted by a gaussian window, indicated by the overlaid circle. Right: The image gradients are added to an orientation histogram. Each histogram include 8 directions indicated by the arrows and is computed from 4x4 sub-regions. The length of each arrow corresponds to the sum of the gradient magnitudes near that direction within the region [58] .......... 49

Figure 3.11  Organization of curvelet transform [64] ................. 52
LIST OF FIGURES AND TABLES

Figure 3.12  An example: Spatial decomposition of single sub-band [64]..53
Figure 3.13  Curvelet tiling of the frequency plane. In frequency domain curvelets are supported near symmetric ‘parabolic’ wedges. The shaded area represents this wedge [65]..........................54

Figure 3.14  Wavelet versus Contourlet: illustrating the successive refinement by the two systems near a smooth contour, this is shown as a thick curve separating two smooth regions [69]........................................................................57

Figure 3.15  LP: One level of decomposition. The outputs are a coarse approximation \( c[n] \) and a difference \( d[n] \) between the original signal and the prediction. \( H \) and \( G \) are called (lowpass) analysis and synthesis filters respectively and \( M \) is a sampling matrix [69].................................................................58

Figure 3.16  DFB. Frequency partitioning where there are \( 2^3 = 8 \) real wedge-shaped frequency bands. Sub-bands 0–3 correspond to the mostly horizontal directions, while sub-bands 4–7 correspond to the mostly vertical directions [69]..............................59

Figure 3.17  Two-dimensional spectrum partition using quincunx filter banks with fan filters. The black regions represent the ideal frequency supports of each filter. \( Q \) is a quincunx sampling matrix[69] ............................................................59

Figure 3.18  Contourlet filter bank. First, a multi-scale decomposition into octave bands by the LP is computed, and then a DFB is applied to each band-pass Channel[69].................................60

Figure 3.19  (a)-(c) represents different linear hyperplanes for two-class data labelled as ‘+1’ and ‘-1’ respectively, (d) various hyperplanes (e) maximal margin hyperplane (i.e. optimal hyperplane)[81] ......................................................62

Figure 3.20  An example of separable problem in a 2 dimensional space. The support vectors, marked with gray squares define the margin of largest separation between two classes[37]........63

Figure 3.21  Example of Feature mapping. (Left) input space (right) feature space[82]..............................................................66

Figure 3.22  The test sample (green circle) should be classified either to the first class of blue squares or to the second class of red triangles. If \( k = 3 \) it is classified to the second class because there are 2 triangles and only 1 square inside the inner circle [95].......70
LIST OF FIGURES AND TABLES

Figure 3.23  Car images taken for preparation of database from Loughborough University ............................................ 77

Figure 3.24  Example Car images taken from the database provided by [23] ............................................................... 78

Figure 4.1  Proposed, 2D-LDA based Vehicle MMR System .................. 83

Figure 5.1  The top figure shows the extracted features for two car images of same make and model (Ford Fiesta). Both cars have features which do not correspond. The bottom figure shows the matches between individual keypoints, when the straightforward SIFT based approach is used. Only two of the displayed features found a match, out of which only one is correct ................. 96

Figure 5.2  Proposed SIFT based approach to vehicle MMR ............... 97

Figure 6.1  Matched pair of keypoints [34] ........................................ 104

Figure 6.2  Proposed System ................................................... 105

Figure 6.3  Examples of the database of images ............................. 106

Figure 6.4  Detected SIFT features of individual images (top row) and projected keypoints of a particular make-model on a sample image ................................................................. 107

Figure 6.5  Two examples (a-c) BMW-3, (d-f) Citroen. Top row image shows all keypoints from the training set of a class projected onto one selected image. Middle row images represent 25% of feature selection based on highest entropy. Bottom row images represent 25% of feature selection based on lowest entropy values ................................................................. 112

Figure 6.6  Feature selection from top to bottom row: no features, all keypoints, 25% highest entropy keypoints, 25% lowest entropy

Automatic Approaches Towards Vehicle Make and Model Recognition XVII
LIST OF FIGURES AND TABLES

keypoints, on example classes from left to right: Audi-A4, Toyota corolla and Citroen........................................114

Figure 6.7 Matching results.................................................................115

Figure 7.1 Multi-scale and multi-directional decomposition of a car image using contourlet transforms...........................................120

Figure 7.2 Proposed Recognition Algorithm............................................122

Figure 7.3 Contourlet decomposition of an image of a car with equal sized scales and directions. The scale levels are represented by the colour blocks.................................................................123

Figure 7.4 SVM classification error vs. size of training class. (a)-(f): directly using Contourlet coefficients from different combinations of scales-levels, (g): using the proposed Contourlet feature selection criteria..................................................129

Figure 7.5 (a): SVM ‘one against all’ classification error vs. training class for using standard deviation of scale-levels 3&4 of [33]...1 (b): SVM ‘one against one and all’ classification error vs. training class for using standard deviation of scale-levels 3&4 of [33].................................................................130

Figure 7.6 Classification accuracy vs. training class size.........................131

Figure 7.7 (a): SVM vs. k-NN classification results on extracted features from section 7.2.3.................................................................133 (b): SVM vs. k-NN classification results on reduced features after dropping 100 projection axis from section 7.2.4.............133

TABLES

Table 3.1 25 classes for various makes and models of cars...............75

Table 3.2 subset of 25 classes from car database of [23]..............76

Table 4.1 Comparison of 2D-LDA vs. PCA under normal lighting conditions using all eigenvectors. Only cases where results defer have been illustrated..........................88
LIST OF FIGURES AND TABLES

Table 4.2 2DLDA vs. PCA under varying lighting and occlusion conditions considering all eigenvectors. Only cases where results defer have been illustrated...............................89

Table 4.3 2DLDA vs. PCA results in a reduced vector space.................90

Table 5.1 Experimental results..........................................................100

Table 6.1 Classification accuracy vs. number of iterations: BMW-3 vs. rest .........................................................................................110

Table 7.1 Evaluation of SVM and k-NN performance against different kernels and value of 'k' using n-fold cross-validation (values of n are 25,10,6,4,2) respectively........................................127
### ABBREVIATIONS AND NOTATIONS

#### ABBREVIATIONS

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>ANPR</td>
<td>Automatic Number Plate Recognition</td>
</tr>
<tr>
<td>MMR</td>
<td>Make and Model Recognition</td>
</tr>
<tr>
<td>ROI</td>
<td>Region of Interest</td>
</tr>
<tr>
<td>LDA</td>
<td>Linear Discriminant Analysis</td>
</tr>
<tr>
<td>2DLDA</td>
<td>Two Dimensional Linear Discriminant Analysis</td>
</tr>
<tr>
<td>SIFT</td>
<td>Scale Invariant Feature Transform</td>
</tr>
<tr>
<td>PCA</td>
<td>Principal Component Analysis</td>
</tr>
<tr>
<td>3D</td>
<td>3 Dimensional</td>
</tr>
<tr>
<td>GMM</td>
<td>Gaussian Mixture Model</td>
</tr>
<tr>
<td>SVM</td>
<td>Support Vector Machine</td>
</tr>
<tr>
<td>k-NN</td>
<td>K-Nearest Neighbor</td>
</tr>
<tr>
<td>LHD</td>
<td>Line Segement Hausdorff Distance</td>
</tr>
<tr>
<td>FLD</td>
<td>Fisher Linear Discriminant</td>
</tr>
<tr>
<td>RANSAC</td>
<td>Random Sampling Consensus</td>
</tr>
<tr>
<td>2DFLDA</td>
<td>Two Dimensional Fisher Linear Discriminant Analysis</td>
</tr>
<tr>
<td>AMD</td>
<td>Assembled Matrix Distance</td>
</tr>
<tr>
<td>ADABOOST</td>
<td>Adaptive Boosting</td>
</tr>
<tr>
<td>PC</td>
<td>Principal Component</td>
</tr>
<tr>
<td>1-D</td>
<td>One Dimensional</td>
</tr>
</tbody>
</table>
### ABBREVIATIONS AND NOTATIONS

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>DOG</td>
<td>DIFFERENCE OF GAUSSIAN</td>
</tr>
<tr>
<td>DCT</td>
<td>DISCRETE CURVELET TRANSFORM</td>
</tr>
<tr>
<td>FFT</td>
<td>FAST FOURIER TRANSFORM</td>
</tr>
<tr>
<td>LP</td>
<td>LAPLACIAN PYRAMID</td>
</tr>
<tr>
<td>DFB</td>
<td>DIRECTIONAL FILTER BANK</td>
</tr>
<tr>
<td>2-D</td>
<td>TWO DIMENSIONAL</td>
</tr>
<tr>
<td>RBF</td>
<td>RADIAL BASIS FUNCTION</td>
</tr>
<tr>
<td>NMC</td>
<td>NEAREST MEAN CLASSIFIER</td>
</tr>
</tbody>
</table>

### NOTATIONS

<table>
<thead>
<tr>
<th>Symbol</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>$C$</td>
<td>COVARIANCE MATRIX</td>
</tr>
<tr>
<td>$I$</td>
<td>IMAGE</td>
</tr>
<tr>
<td>$P_i$</td>
<td>ITH PRINCIPAL COMPONENT</td>
</tr>
<tr>
<td>$\Gamma$</td>
<td>IMAGE IN VECTOR FORM</td>
</tr>
<tr>
<td>$\Psi$</td>
<td>MEAN IMAGE</td>
</tr>
<tr>
<td>$\Omega$</td>
<td>IMAGE IN EIGENSPACE</td>
</tr>
<tr>
<td>$H$</td>
<td>CLASS</td>
</tr>
<tr>
<td>$S_B$</td>
<td>BETWEEN CLASS SCATTER MATRIX</td>
</tr>
<tr>
<td>$S_W$</td>
<td>WITHIN CLASS SCATTER MATRIX</td>
</tr>
<tr>
<td>$Z_i$</td>
<td>SIZE OF CLASS $i$</td>
</tr>
<tr>
<td>$W_{opt}$</td>
<td>OPTIMAL PROJECTION</td>
</tr>
<tr>
<td>$h$</td>
<td>NUMBER OF CLASSES</td>
</tr>
<tr>
<td>$E$</td>
<td>TRAINING SET SIZE</td>
</tr>
</tbody>
</table>
ABBREVIATIONS AND NOTATIONS

<table>
<thead>
<tr>
<th>Symbol</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \alpha )</td>
<td>CONSTANT</td>
</tr>
<tr>
<td>( \beta )</td>
<td>CONSTANT</td>
</tr>
<tr>
<td>( \Phi )</td>
<td>MEAN SUBTRACTED IMAGE</td>
</tr>
<tr>
<td>( \Delta )</td>
<td>BAND PASS FILTER</td>
</tr>
<tr>
<td>( P_o )</td>
<td>LOW PASS FILTER</td>
</tr>
<tr>
<td>( l )</td>
<td>CLASS LABEL</td>
</tr>
<tr>
<td>( \delta )</td>
<td>CONSTANT</td>
</tr>
<tr>
<td>( D )</td>
<td>DISTANCE MEASURE</td>
</tr>
<tr>
<td>( G(x,y,\sigma) )</td>
<td>GAUSSIAN FUNCTION</td>
</tr>
<tr>
<td>( \text{Diff}(x,y,\sigma) )</td>
<td>DIFFERENCE OF GAUSSIAN</td>
</tr>
<tr>
<td>( \hat{h} )</td>
<td>WEAK HYPOTHESIS</td>
</tr>
<tr>
<td>( \text{Dis} )</td>
<td>WEIGHT DISTRIBUTION</td>
</tr>
<tr>
<td>( R )</td>
<td>NUMBER OF ITERATIONS</td>
</tr>
<tr>
<td>( \hat{H} )</td>
<td>FINAL HYPOTHESIS</td>
</tr>
<tr>
<td>( \hat{D} )</td>
<td>TOTAL NUMBER OF DIRECTIONS</td>
</tr>
<tr>
<td>( \hat{L} )</td>
<td>TOTAL NUMBER OF SCALE LEVELS</td>
</tr>
<tr>
<td>( \hat{d} )</td>
<td>SINGLE DIRECTION</td>
</tr>
<tr>
<td>( \text{Map}_d )</td>
<td>DIRECTIONAL MAP</td>
</tr>
<tr>
<td>( \hat{C}_{i+1} )</td>
<td>COARSE LEVEL IMAGE</td>
</tr>
<tr>
<td>( \text{Tex} )</td>
<td>LOCAL TEXTURE</td>
</tr>
<tr>
<td>( \text{Ed} )</td>
<td>ORIENTED EDGE</td>
</tr>
<tr>
<td>( \hat{p} )</td>
<td>NUMBER OF SUBINTERVALS</td>
</tr>
<tr>
<td>( L )</td>
<td>NUMBER OF CLASS LABELS</td>
</tr>
</tbody>
</table>
CHAPTER 1

An Overview

1.0 INTRODUCTION

The need for public security has substantially increased in recent years. This has resulted in the development of a large number of computer vision systems that have been successful in exploiting the strengths of current technical developments, in providing automatic/semi-automatic, real-time, video surveillance of public places.

To this effect, access control systems have been implemented in parking areas, buildings and other restricted sites. In particular, access control systems based on recognition and verification of vehicles using number-plates, i.e. Automatic Number Plate Recognition (ANPR), is most common. An extensive amount of research has been performed in the area of ANPR and state-of-art has been developed.

Traffic monitoring and toll collection systems are another example for the application of vehicle recognition. To this effect a significant amount of research has been carried out in the area of computer vision based vehicle classification. However these classification techniques have been limited mostly to algorithms distinguishing between different categories of vehicles i.e. car, bus, truck etc.

In contrast, an effective vehicle recognizing system solicits the need of correctly identifying the make and model of vehicles within a given category. Several vehicle recognition systems based on correctly recognizing only the vehicle registration plates, are in widespread commercial use at present. However reports by police and media sources have indicated that number-plate cloning, i.e. using bogus registration plates, has been recently used to breach the security. This has been used to break security provided by Automatic Number Plate Recognition (ANPR)
CHAPTER 1

techniques at automatic number plate identification based access control systems and avert congestion charges in busy city areas. This problem can be addressed by enhancing the reliability of access control systems by using a hybrid approach, i.e. the combined use of ANPR and Vehicle Make & Model Recognition (MMR), automatic identification of a vehicle’s visual description comprising of either one or more of properties such as, make, model and colour techniques.

1.1 RESEARCH MOTIVATION

Vehicle MMR is a comparatively new research area. A relatively limited number of techniques that directly relate to vehicle MMR have been proposed in the literature. Petrovic and Cootes [23] proposed techniques for the recognition of cars, by extracting gradient features from images. A number of feature extraction algorithms including direct and statistical mapping methods were applied to Regions-Of-Interest (ROIs) of frontal views of cars, to obtain sampled structures. These feature vectors were then extracted and classified using simple nearest neighbor classification methods. Daniel T.Munroe and Michael G.Madden [28] investigated the use of machine learning classification techniques in vehicle MMR. Initially a Canny edge detector followed by a dilation process was used to extract feature vectors. Subsequently different machine learning classifiers were used to determine vehicle make and model associated with each feature vector. L. Dlagnekov [25] explored the problem of MMR by using Scale Invariant Feature Transforms (SIFT) [58]. It is used to identify the points of interest in car images which are subsequently utilized in matching. Michal Conos [27] proposed a modified simple version of SIFT descriptors [58]. David Anthony [31] extended the work done by Dlagnekov [25] by replacing the SIFT features with features which characterize contour lines. In this approach, initially, edges are extracted from the rear views of car images. These are then extended to line segments by using a strip based line generator algorithm and are subsequently used in matching. Kazemi et. al. [32], investigated the use of Fast Fourier Transforms, Discrete Wavelet Transforms and Discrete
Curvelet Transforms [66] based image features in vehicle MMR. Pablo et.al in [29] proposed an oriented-contour point based voting algorithm for multiclass vehicle type recognition, which was particularly proved to be effective under occlusion. S.Rahati et.al [33] proposed the direct replacement of Curvelet transforms with Contourlet transforms [69] in Kazemi et. al.'s [32] proposal for vehicle MMR. Hyo Jong Lee [26] proposed the use of texture descriptors: contrast, homogeneity, entropy and momentum for the identifications of vehicles. S.Cheung and A.chu [34] improved the work of Dlagnekov[25] by introducing better keypoint matching and methods to prune the outliers.

A number of different approaches have been published in the literature for vehicle MMR as discussed above, the search for a robust, efficient algorithm still remains an open research problem. All vehicle MMR approaches proposed in the literature, summarized above, are based on an initial stage of feature detection, where the detected features are subsequently used in matching. For example, the majority of the methods are based upon using edge maps for global feature extractions. However, the pixel domain edge extractors used (e.g. Canny [108]) are limited in their performance and therefore fail in accurately capturing the smooth curves/contours which are an important part of most of the images. Kazemi et al's proposal of [32] which is a vehicle MMR method based on Discrete Curvelet transforms [61, 65] is the first attempt to address this problem. Curvelet transforms provide a multiresolution, band pass and directional functional analysis method, which is useful to represent the image edges and curved singularities more efficiently. However a major challenge in capturing the geometry and the directionality in images comes from the discrete nature of the data. The Curvelet transform, that is initially developed in the continuous domain and later discretized for sampled data is not effective to be used with digital images. The use of SIFT in vehicle MMR by Louka[25], a local feature extraction approach, promises to address some shortcomings of traditional global feature based approaches by providing interest points invariant to scale, rotation and illumination. However, the matching strategy provided in his paper needs further improvements.
The shortcomings of the existing approaches to vehicle MMR motivated the research work presented in this thesis. One main focus of the research is to investigate the limitations of the existing techniques and to propose robust feature extractors, which are combinations of local and/or global feature extractors. Literature on face recognition methods (see chapter 2), has proved that appearance based methods such as Two Dimensional Linear Discriminant Analysis (2DLDA) [55, 56] performs well under illumination variations. This provided motivation to explore the use of appearance based techniques in vehicle MMR, where illumination variations provide practical challenges. Further though SIFT provides features which are invariant to scale, rotation and other variations, these features largely depend on distinctive regions such as blobs and well textured patches. In the application domain of vehicle MMR, cars are textureless and inter-class differences are very small. Therefore using SIFT [58] as the only reliable features for recognition purposes may not provide the required discrimination accuracy. The need is to improve the description of an object through the use of additional shape features.

1.2 AIM AND OBJECTIVES

Aim: To design and develop robust, efficient algorithms for the detection and classification of vehicle make and model, giving rise to systems that will be able to enhance the fool proof nature of existing vehicle access systems.

The specific research objectives can be listed below as:

- To study possible shortcomings of existing techniques and propose novel ideas towards their improvement.
- To investigate the possible application of 2DLDA in vehicle MMR. Further, effects of selectively using eigenvectors needs to be investigated to check the robustness of 2DLDA under varying illumination conditions and partial occlusion.
• To propose novel approaches towards SIFT feature selection and to develop feature matching algorithms that can significantly improve vehicle MMR accuracy and robustness.

• To propose the use of specific local texture features in the lowpass subband and the oriented image edge features across the scales of directional resolutions of Contourlet transformed image as features as an alternative to the direct use of Contourlet transform coefficients as features.

• To select the most distinctive and representative SIFT features of a particular class for matching by the use of the AdaBoost algorithm.

• To identify the limitations of the proposed algorithms and outline the future directions of research.

1.3 CONTRIBUTIONS OF RESEARCH

The following original contributions have been made by the research presented in this thesis (see Chapters 4-7, inclusive). The full publication details of the corresponding conference and journal papers have been included in the Appendix-A, referred to as A1 to A7.

1.3.1 Use of Two Dimensional Statistical Linear Discriminant Analysis (2DLDA) in Vehicle MMR

In chapter 4, a real time, appearance based, vehicle MMR approach using 2DLDA [55, 56] has been proposed that is particularly capable of addressing the problems of illumination variation and partial occlusion. Principle Component Analysis (PCA) [49] has been used as a benchmark to compare and contrast the performance of the proposed 2DLDA approach to vehicle MMR. Experimental results have been
provided to analyse the proposed algorithm's robustness under varying illumination and occlusions conditions. In addition, detail analysis of performance of 2DLDA by dropping varying significance eigenvectors has been investigated. It has been shown that the best performance with the proposed 2DLDA based vehicle MMR approach is obtained when the eigenvectors of lower significance are ignored. For the given database of 200 car images of 25 different make-model classifications, a best accuracy of 91% was obtained with the 2DLDA approach. It has been concluded that in general the 2DLDA based algorithm supersedes the performance of the PCA based approach.

1.3.2 Use of Scale Invariant Feature Transforms (SIFT) in Vehicle MMR

The focus of chapter 5 is to improve the matching scheme of the popular SIFT approach [58]. Instead of performing a keypoint to keypoint match from test image to each training image, each keypoint from the training image is matched to a window of pixel descriptors around the corresponding location on the test image. Experimental results have been provided to prove its effectiveness in vehicle MMR. It has been shown that the proposed technique is capable of identifying Vehicle MMR under image scale, blur variations and within practical limits of rotation, translation and occlusion.

1.3.3 Use of Adaptive Boosting in Feature Selection for Vehicle MMR

In chapter 6 a novel approach to SIFT based vehicle MMR is proposed in which SIFT features are initially investigated for their relevance in representing the uniqueness of the make and model of a given vehicle class based on Adaptive Boosting (AdaBoost) [97, 98]. Experimental results are provided to show that the proposed selection of SIFT features significantly reduces the computational cost associated with classification at negligible loss of the system accuracy. It is further proved that the use of more appropriate vehicle matching algorithms enables
significant gains in the accuracy of the algorithm. Experimental results are provided to prove that a 92% accuracy rate can be achieved by using the proposed SIFT based matching scheme in chapter 5, on a publically available database of car frontal views.

1.3.4 Use of Contourlet Domain Localised Directional Feature Maps in Vehicle MMR

The usefulness of multi-resolution based feature analysis techniques leading to efficient object classification algorithms has received close attention from the research community. To this effect, Contourlet transforms [69] that can provide an efficient directional multi-resolution image representation has recently been introduced. In chapter 7, a novel localized feature detection method in Contourlet transform domain has been proposed that is capable of increasing the classification rates up to 4%, as compared to the previously proposed Contourlet based vehicle MMR approach [33] in which the features are non-localized and thus results in sub-optimal classification. Further it has been shown that the proposed algorithm can achieve the increased classification accuracy at significantly lower computational complexity due to the exclusive use of directional sub-bands most significant to images consisting of vehicles.

1.4 ORGANISATION OF THESIS

For clarity of presentation the thesis has been organized into eight chapters as described below.

Chapter 1 provides an overview of the research domain identifies open research problems and provides a summary of the proposed solutions. The chapter provides
CHAPTER 1

the research aim, identifies the specific objectives and clearly states the original contributions made by the research presented in the thesis.

Chapter 2 presents details of the existing literature on vehicle MMR, car identification and techniques adopted from face recognition and analyses their performance. The chapter further discusses in detail, different vehicle MMR methods and identifies the advantages and disadvantages of each method.

Chapter 3 concentrates on providing the background knowledge related to the research context in which the novel vehicle MMR algorithms have been proposed. In particular the chapter details Two Dimensional Linear Discriminant Analysis (2DLDA)[55,56], Scale Invariant Feature Transform (SIFT)[58], Contourlet Transform[69] and Adaptive Boosting(AdaBoost)[97,98] which are used in this thesis to identify visually significant and discriminant features.

Chapters 4, 5, 6 and 7 provide information on original contributions made by the work presented in this thesis to the field of vehicle MMR. The chapters specifically include the novel methodology adopted, experimental design details, results, analysis and suggestions for further improvement.

Finally Chapter 8 concludes the research presented in this thesis with an insight into the future directions of research.
CHAPTER 2

Literature Review: Vehicle Make and Model Recognition (MMR) Techniques

2.0 INTRODUCTION

This chapter provides an insight into existing research on vehicle MMR. In addition, a number of state-of-the-art algorithms in vehicle MMR and/or face recognition, which provide inspiration to the novel algorithms proposed in the thesis and thus used as benchmarks for evaluating their performance, have been discussed in detail. Further a brief overview of key algorithms in detection of cars in images/scenes and details of preparing the database of images used in the experimentation have been included.

This chapter is divided into several sections for clarity of presentation. Section 2.1 provides a review of literature on the detection of cars from a given scene. These algorithms provide a viable pre-processing stage to vehicle MMR and are thus considered important in providing an end-to-end solution to vehicle MMR in practice. Section 2.2 provides a detailed literature review on vehicle MMR. Section 2.3 provides a number of state-of-the-art techniques used in face recognition which have provided some inspirations to the novel algorithms proposed in this thesis. Finally section 2.4 summarises, making conclusions.
CHAPTER 2

2.1 DETECTION OF CARS IN SCENES

Detection of vehicles, in particular cars, from images containing natural scenes is the first step towards vehicle MMR. Extensive research has been performed and a large amount of literature has been published in the detection of vehicles in the video sequences/images and their subsequent classification into various categories/types such as cars, buses, trucks [1-16] etc. These systems have been used in analysing road usage and highway design/maintenance. Although a complete review of all existing literature in this area of research is beyond the scope of this thesis, details of a number of key approaches have been provided below. Note that detection of cars in a scene is the first stage in any end to end solution to vehicle MMR.

C. Papageorgiou and T. Poggio [17]

This paper focuses on developing a general, trainable architecture for the detection of cars that has previously been applied in the detection of human faces and people in static images. In this technique a model for a car is learned using a set of labelled training data of cars. The training images are first transformed from the pixel space to that of Haar wavelets to obtain a compact representation. The feature vectors are then used to train a support vector machine classifier. The performance of the algorithm is presented with the help of an ROC curve. The detection rate is claimed to be 90%.

H. Schneiderman and T. Kanade [18]

This paper describes a statistical method for 3D object detection. The main contribution is detecting faces with out-of-plane rotation and cars over a wide range of viewpoints. The statistics of both object appearance and “non-object” appearance is represented using a product of histograms. Each histogram represents the joint statistics of a subset of wavelet coefficients and their position on the object. Many
CHAPTER 2

such histograms are used to represent a wide variety of visual attributes. The best achieved accuracy is reported to be 92%.

G. Dorčo and C. Schmid [19]

This paper introduces a novel method for extracting discriminant, scale-invariant feature points, constructing rotation invariant descriptors for these feature points and clustering these descriptors into object parts. Once object parts are obtained, feature selection techniques: likelihood ratio and mutual information are used to determine the most discriminant ones. A classifier is then learned for each of these parts. Support Vector Machines (SVM) and Gaussian Mixture Models (GMM) have been used as classifiers. The advantage of this technique is that it provides robust part detection, and it is scale, rotation and illumination invariant. The proposed method is evaluated in car detection tasks with significant variations in viewing conditions, and promising results have been demonstrated.

Z. Zhu, H. Lu, J. Hu and K. Uchimura [20]

This paper presents a car detection technique based on multi-cues. At the bottom level, two area templates based on edge cues and interest points cues, are first designed, which can reject most of non-car sub-windows (i.e. areas from the background). At the top level, both global and local texture cues are considered. To characterise the global structure property odd Gabor moments are introduced and are trained by a Support Vector Machine (SVM) [82]. The final experiment results show that the integration of global structure property and local texture property is more powerful in discrimination between car and non-car objects. A detection rate of 93% has been reported.
This paper presents the task of building a component based car detection technique in natural images of street scenes. The intuitive motivation behind this idea is that each part of an object should be less sensitive to changes in illumination, pose, and rotation than the object as a whole. First of all, an interest operator (Scale Invariant Feature Transform (SIFT)) is applied to extract the key-points in the training data and test image. Car-specific key-points are learned for the training data by using clustering. Key-points extracted from the test image are then compared against these car-specific key-points and a similarity vector is obtained. This vector is input to a Support Vector Machine (SVM) for classification.

Similar to B. Leung [21], this paper also focuses on the detection of cars based on components. A vocabulary of car parts, such as tyres and windshields etc., is automatically created from training images and is subsequently used as part representations of the car images. Detection of cars is done by finding individual parts and comparing their spatial similarity. The algorithm has been evaluated on difficult sets of real-world images containing side views of cars, and is seen to successfully detect objects with background clutter and mild occlusions.

2.2 VEHICLE MMR ALGORITHMS

Vehicle MMR is a comparatively new research area. A comprehensive literature survey was carried out which revealed that only a relatively limited number of techniques that directly related to this subject area, exists.

Generally, vehicle MMR undergoes the following key stages:

- Region of Interest(ROI) Detection
CHAPTER 2

• Feature Detection, Extraction and Classification

The literature on vehicle MMR can be categorised under the above two areas and presented as follows:

2.2.1 Region of Interest (ROI) Detection

The recognition task begins with finding out the regions in car images which contain distinguishable features. From the literature survey and the subjective tests performed [117] on how people identify the specific model of a car, it has been concluded that the front or rear views of cars, which include the regions around the number-plate (grille, badge, lights), are sufficient to extract the required features. Some techniques for ROI cropping, rely on first locating a reference segment on the car, in this case the frontal/rear number plate.

V. Petrovic and T. Cootes in [23, 24] find all possible right-angle corners using suitably tuned, separable gradient filters from the front-end view of cars. A hierarchical algorithm for aggregation of corner points into valid rectangular constellations is used to generate a hypothesis for the location of the number plate in the image. The location and scale of this detected segment is used as a reference to define regions-of-interest (ROI). ROI is defined in terms of number-plate width \( w_p \) relative to its centre.

L Dlagnekov in [25] crops a fixed window (i.e. ROI) around the license plate of the middle frame of each tracked sequence of video using a novel license plate detector [25]. The cropped window is positioned such that the license plate is centred in the bottom third of the image. The ROI is collected from the rear end of the car images.
CHAPTER 2

In H.J.Lee [26], the ROI is defined around the upper front region of vehicle including the radiator grille and part of the head lights. In order to find the ROI, a base line has been set at the center of the license plate. A bounding box is tweaked on the base line from the left to the right to adjust the best position. The license plate is detected by investigating both geometric information of license plate and intensity changes between numbers and background [26].

In M. Conos [27], after correctly detecting a number-plate from the front view of a car, two corners of the number-plate: top-left and bottom-right are used to establish an orthogonal geometry. With the top-left corner of the number-plate as the origin a region surrounding the number-plate is cut using the width ‘w’ and height ‘h’ of the number-plate. ROI for [23] and [27] is depicted in Figure 2.1.

![Figure 2.1: Car Geometry measured within the space of the number-plate (a) [23], (b) [27].](image)

In D.T.Munroe [28] an approach that doesn’t rely on the license plate as the reference point for ROI extraction is proposed. Instead, the top half of the front view of the whole car image is assumed to be the ROI. This approach suffers from the fact that the ROI may not be aligned spatially even for images containing cars that belong to the same make and model.
P.Negri et al in [29] uses the License Plate Recognition Editor’s (LPREditor) license plate recognition system [30] for obtaining four corner points of a vehicle license plate. The ROI is subsequently extracted as a rectangular window, following a procedure similar to that used in [23] and [27].

2.2.2 Feature Detection, Extraction and Classification

These form the key stages of a vehicle MMR system after the images have been captured and the ROIs have been cropped. Following is a review of literature of the state-of-the-art.

V.Petrovic and T.Cootes [23]

This paper focuses on the recognition of rigid structure samples obtained using specific feature extraction techniques. A number of feature extraction algorithms, including direct mapping (soble edge response, direct normalized gradients, locally normalized gradients, square mapped gradients, Harris corner response and spectrum phase) and statistical mapping (principal component analysis) are applied to extracted ROIs (as explained in section 2.2.1), to obtain sampled structures. Gradients are normalized between \((-\pi, +\pi)\). These feature vectors are then classified using simple nearest neighbour classification methods. Experiments have been performed on a database of front views of car images with 1132 images ordered into 77 distinct classes. Out of these, 105 have been used for training and 1027 for testing purposes. Images of the cars represent a range of weather and lighting conditions. It was observed that among all feature extractors, square mapped gradient performs best. It is insensitive to small changes in gradient orientation, noise and also helps in maintaining the global and local structure of objects in a robust manner. Accuracy obtained with this extractor is 97.7%.
CHAPTER 2

V.Petrovic and T.Cootes [24]

The previous recognition system [23] is limited by its capability in handling the pose variation of rigid samples. Rigid samples are sensitive to pose variation and thus confusions can be created due to large number of potential classes. Petrovic [24] provides a refinement by using a novel match refinement algorithm that maximises the discrimination between the subset of most likely classes by optimising for object pose and adaptively normalising feature vectors. The refinement algorithm thus increases the robustness by making adjustments using optimization techniques such as the simplex algorithm [116]. Identification accuracies of up to 94.4% have been reported.

D.Munroe and M.Madden [28]

The paper investigates the use of machine learning classification techniques in vehicle MMR. A Canny edge detector followed by a dilation process is used to extract a fixed length feature vector from the ROI (see section 2.2.1). Subsequently different machine learning classifiers including C4.5 decision tree, the k-Nearest Neighbour (k-NN)[95](see chapter 3) and a feed-forward neural network trained using back propagation are used to determine the vehicle make and model associated with each feature vector. The algorithm has been tested on a dataset of 150 frontal view car images (30 images of each of five classes). Two sets of experiments have been performed. The first one involves the use of multi-class classifiers: C4.5 decision tree, k-NN and a feed-forward neural network trained using back propagation. An evaluation of the performance has shown that the Neural Network and the k-NN classifier perform better than C4.5 decision tree algorithm. Classification accuracies of up to 97% have been achieved.
L. Dlagnekov [25]

In this paper, a vehicle MMR approach which uses the Scale Invariant Feature Transform (SIFT) [58] (see chapter 3) as the feature extraction mechanism, has been proposed. SIFT is invariant to scale, rotation and partially to illumination differences [58]. The algorithm starts by finding out SIFT keypoints and their descriptors in the ROI (see section 2.2.1) of the test and training image sets. Matching of the test image to each of the training images is performed subsequently. This is achieved by finding the smallest $L_2$ distance of each keypoint in the test image to that of any training image. A threshold is set as an upper bound for the distance value. The training image which gives the largest count of matched keypoints is considered as the best match. A further pruning procedure is applied to get rid of keypoints which contribute to false matches (e.g. keypoints in number-plate region). To test the algorithm, a database of 1,140 car images is generated. The SIFT matching algorithm yields a recognition rate of 89.5% on the query set. This technique has an advantage as it doesn’t depend upon the proper registration of the images to match.

D. Anthony [31]

In this paper, David [31] extends the work done by Dlagnekov [25]. Though SIFT features detect interest points which provide robust recognition systems; they only focus on minute portions of images. Discriminating global features such as shapes and patterns remain ignored. David improves this by replacing SIFT features with features which characterize contour lines. ROIs [25] (see section 2.2.1) from the back-ends of cars are collected. Edges are then extracted from the images and are termed as edge maps. These edges are then extended to line segments, which are partially invariant to illumination changes and capture local structure information at different scales, which is not possible when using SIFT. Further line segments are generated using a strip based algorithm. Line segments are then compared using a line Segment Hausdorff distance (LHD) metric [35]. The vehicle MMR task is
evaluated using a database of 1103 cropped images of the back end of vehicles of various makes and models. A further set of 38 query images are used for testing. The reported overall recognition rate of the system is 59.7%. This technique is good at recognizing images based on their discriminating global features like shapes and patterns, which can be considered as forming the distinct features of most vehicles. Line segments are partially invariant to illumination change, which is advantageous in the practical application domain of vehicle MMR. However, the computation of the Line Segment Hausdorff distance is complex and thus the overall identification speed is slow. Further, only a limited number of images can be used as a model database, which affects overall recognition rate of the system.

M.Conos [27]

The work of M.Conos [27] focuses on the extraction of SIFT descriptors, which are a modified and simplified version of the original SIFT descriptors [58] (See chapter 3). The process starts with processing the ROI (see section 2.2.1) by feature extractors. Instead of finding keypoints using the traditional SIFT approach, every pixel in the ROI is considered a keypoint. The feature extraction initiates by first dividing the image into square patches. The gradient magnitude and orientation is calculated for each pixel in the image. Every square patch is then divided into a certain number of bins. Every bin represents a certain interval of orientation which is assigned to it. The values assigned to bins are calculated using the 3 SIFT variants: SIFT-Ori, SIFT-Grad and SIFT-Grad Wei. Each SIFT representation has either overlapping tiles or non-overlapping tiles. Subsequently a feature vector is constructed by putting together all the bins within all tiles. Using k-NN and Fisher Linear Discriminant analysis (FLD) as the classifier, vehicles are finally classified into different classes. The database used in this case consists of 250 images of cars. From the experimental results it is concluded that the overall classification rate is 94%. The classification accuracy can be improved by increasing the number of tiles. Best results are achieved when using the SIFT topology (number of square patches
in the horizontal direction, number of square patches in the vertical direction and number of bins per square patch) 15*9*15 and overlapping tiles.

P. Negri et al. [29]

In this paper, a model based system for vehicle type recognition and classification is proposed. The focus is to recognise cars under occlusion conditions. A vehicle class is represented by an Oriented-Contour Points based model. During the initial phase, using the training samples, an Oriented-Contour based Model for every vehicle type class of the set of available classes, called Knowledge Base is produced. The classifier is based on a voting algorithm and on a Euclidean edge distance. When a new image is presented as an input, a scoring function is computed for every available class in the Knowledge Base. Input that is identified as the best match from the Knowledge Base is considered to be the class with the highest score. The experiments involve a finite set of $k = 20$ classes named a confusion matrix. It has been reported that the proposed Oriented-Contour Voting algorithm correctly identifies in average 92.39% samples from the confusion matrix.

H. Lee [26]

In this paper a novel method is proposed to identify vehicles based on specific information, i.e., colour, license plate, and vehicle's model. The ROI is detected using the technique presented in [26] (See section 2.2.1). Initially a gray level co-occurrence matrix is calculated from the front region of the vehicles. Texture descriptors are then calculated using four texture descriptors: contrast, homogeneity, entropy and momentum. A good separator distinguishes one model from all others. However, some descriptors are not distinguishable from two different models. Thus, four texture descriptors contrast, homogeneity, entropy and momentum are computed in four directions east, west, south-east and north-west and used all together. These 16 texture descriptors are then fed into to a three layer neural network. The proposed algorithm has been tested on 24 different models of 415
vehicles. The performance has been measured with mixed data of trained and untrained data sets giving accuracy rates of up to 94%.

F.Kazemi et.al [32]

This paper provides a comparative study of the application of three different kinds of feature extractors on car images for recognition purposes. The feature extractors used include Fast Fourier transforms, Discrete Wavelet transforms and Discrete Curvelet transforms [66] (see chapter 3). Extracted features are then fed into a k-NN classifier for classification. Further, the dimension of feature vectors is reduced and best features are selected using an inter-class variance to intra-class variance criteria. Intra-class variance translates to variability of the feature within the target class, and a small value would indicate better representation of the feature for this class. On the other hand, inter-class variance measures the separability of different target classes using this feature, and hence, a large value is desirable. To check the performance of the algorithm, a dataset of 300 vehicle rear view images from 5 different classes of vehicles has been collected. Accuracy rates approaching 100% has been achieved by using Curvelet transform (using all coefficients). The comparison of the three proposed approaches for identifying the kind of vehicles shows that the Curvelet transform can extract better features as compared to the other descriptors.

S.Rahati et.al [33]

This recognition system is based on performing a feature extraction of an image by applying Contourlet transform [69] and subsequently finding the standard deviations of each Contourlet sub-band. Three different types of classifiers have been applied subsequently on the feature set. They include, SVM (both one versus one and one versus all) [82] and k-NN classifiers [95]. Experimental results have shown that up to 99% accuracy in recognition can be achieved when using the
standard deviations of sub-bands 3 and 4 as features and the SVM (one versus one) as the classifier.

S.Cheung and A.chu [34]

The work in this paper builds on that of [25]. In the technique proposed, there are three primary steps. The first step consists of detecting interesting features, or interest points using two approaches: SIFT and Harris corner detection, performed on the image of the query car and on the images in the dataset. In the second step of the algorithm, interest points of the query image are compared to sets of interest points of each of the images in the database, based on appearance. Two methods have been investigated for Interest point matching: traditional Lowe's SIFT matching [58] (See chapter 3) and Fast Normalized Cross Correlation [110]. In the final method, the paired interest points are used to find a subset of inliers which fit best to a given geometric transformation model. The outliers are not considered in the matching. Inliers extraction is performed by exploring methods which include, Contour segmentation, mirror symmetry and RANdom SAmpling Consensus (RANSAC) [109] using the affine transformation as the model. Experiments have been performed on different combinations of interest point detectors and interest point matching schemes. In the dataset, the image of an unknown car model can be of a specific view (i.e. rear-view, side view, ¾ view, etc.). The training dataset consists of images taken from the existing database of cars arranged by Dlagnekov [25] along with images taken from websites of various car companies with the same point of view of the car as the query image. The images taken from Dlagnekov’s dataset are primarily composed of rear views of cars whereas the images taken from websites are available in four profile angles: front, rear, side and ¾ views.

A.Ferencz et al [36]

Ferencz et al. was interested in determining whether two images taken at different times and camera orientations are of the same car, where there is really only a single
example that serves as a model. The problem is solved by automatically finding good features on side views of cars from several hundred pairs of training examples, where good features refer to features that are good at discriminating between cars from many small classes. A patch based representation is used, where the distributions of comparison metrics defined on the patches are modelled. Finally, an on-line algorithm is defined that selects the most salient patches based on a mutual information criterion that achieves good performance while only matching a few patches.

2.3 FACE RECOGNITION: LITERATURE

The recognition problem in cars is very much similar to that of face recognition. Both have symmetric geometrical structure. Affixing a sticker on the body or a car or making alterations to the grill or lights of cars is similar to a person wearing accessories (for example spectacles, earrings, and scarf) or changing emotions. However, a number of differences can also be identified. One difference is the large depth of field in car images (for example the distance from the front of the bumper to the front body parts can be 40 cm) compared to faces (distance from the tip of the nose to ear will be generally less than 15 cm). Therefore, in general, the perspective distortions in car images are much larger than in face images. Secondly, faces generally do not have surface reflections whereas shiny body parts of cars can easily reflect bright light. Thus illumination problems due to reflections from lights and body parts of cars can be severe. Car images belonging to the same make and model can have differences in appearance as they relate to different cars, unlike multiple images of the same person. Skin colour of multiple instances of a single person generally remains the same (assuming lighting is maintained constant), whereas cars of the same make and model may be of different colour. The above differences make the problem of vehicle MMR a more challenging problem to solve as compared to face recognition, under controlled conditions of facial expression.
However changes to facial images due to facial expressions can make face recognition a harder problem to solve.

This section reviews a number of face recognition approaches that have provided the basic motivation to achieving the objectives of the research presented in this thesis.

H. Xiong, M. Swamy, M. Ahmad [37]

This paper presented a face recognition approach based on two-dimensional fisher linear discriminant (2DFLD/2DLDA) [54, 55] (see chapter 3) analysis. Experiments have been performed on two face image databases, i.e. on the ORL and UMIST databases. The approach has been compared to face recognition approaches that use PCA [49, 50], PCA+FLD, simple LDA [51] and 2DPCA. It has been proved through experimental results that 2DFLD outperforms all others feature analysis approaches.

Y. Jin, Q. Ruan [38]

In this paper, a novel approach towards measuring the distance between two 2DLDA feature matrices is presented. The approach is named as Assembled Matrix Distance (AMD) [39]. The ORL face database has been used to test the effectiveness of the metric. Results indicate that the AMD metric based 2DLDA method outperforms the traditional 2DLDA method, i.e. achieves higher classification accuracy.

M. Bicego et al [40]

This paper investigates the application of the SIFT approach by D. Lowe [58] in the context of face authentication. In order to determine the real potential and applicability of the method, different matching schemes have been proposed and
evaluated. The matching schemes between the test and training set adopted are different to the original SIFT matching scheme and includes; minimum pair distance, matching eyes and mouth, and matching on a regular grid. The system has been tested on the BANCA database, resulting in producing promising results.

J.Luo et al [41]

This paper proposes the use of person-specific SIFT features, their clustering and a simple matching strategy combined with local and global similarity on these clusters for face recognition. K-means clustering [42] has been used to cluster the image into sub-regions based on the location of features/keypoints. Experiments have been performed on FERET and CASPEAL face databases using only one training sample per person. Results have been compared to reported results of other features such as Gabor wavelet feature and Local Binary Pattern feature. The experimental results demonstrate the robustness of SIFT features to expression, accessory and pose variations.

J Zhang et al [43]

This paper presents the application of Curvelet Transforms [61] (see chapter 3) to the task of face recognition. From the theory of Curvelet [61] it is known that it is a multi-resolution, multi-scale and multidirectional geometric analysis tool which directly takes edges as the basic representation elements and is anisotropic with strong direction. These features of Curvelets have been exploited by taking Curvelet coefficients as a representation of the main features for faces. A support vector machine (SVM) [82] (see chapter 3) has been subsequently used to classify the images. Note that a Multi-class SVM has been used in this paper. The method has been compared with a similar wavelet based method, indicating improved performance.
B. Heisele et al. [44]

In this paper a component-based method and two global methods for face recognition have been presented. The focus of this paper is to evaluate the system performance against varying pose. In the component-based system, facial components are located, extracted and combined into a single feature vector which is classified by a Support Vector Machine (SVM) [82]. The two global systems recognize faces by classifying a single feature vector consisting of the gray values of the whole face image. In the first global system a single SVM classifier is trained for each person in the database. The second system consists of sets of viewpoint-specific SVM classifiers and involves clustering during training. The tests have been performed on a database which includes faces rotated up to about 40 degrees in depth. Results indicate that the overall component-based system outperforms both global systems.

M. Jones, P. Viola [45]

In this paper a new method is presented for face recognition which learns a face similarity measure from examples of image pairs. The features computed are named as 'rectangle' features, which are proved to provide efficient representations. The features compare regions within the input images at different locations, scales, and orientations. Feature selection is done afterwards using an AdaBoost algorithm [97, 98] (See chapter 3) which is trained upon the face similarity function. The weighting procedure in AdaBoost is used to perform feature selection and subsequent classification. Experiments have been performed on the FERET facial database to prove that reducing the feature set can result in better recognition accuracy.
This paper focuses on reducing the dimensionality of feature set and the selection of the most discriminant feature out of a large number of high dimensional Gabor features. Thus, a face recognition system based on AdaBoosted [97, 98] Gabor features has been proposed in this paper. The main contribution of the paper is: (1) AdaBoost is successfully applied to face recognition by introducing the intra-face and extra-face difference space in the Gabor feature space; (2) An appropriate resampling scheme is adopted to deal with the imbalance between the amount of the positive samples and that of the negative samples. The experiments have been performed on the FERET database. The results demonstrate the effectiveness of algorithm when reducing and selecting the number of features. With 700 selected features, 95.2% accuracy rate has been achieved.

2.4 SUMMARY AND CONCLUSION

In this chapter, existing algorithms in vehicle MMR has been introduced and critically analysed. In addition, key algorithms in car detection and relevant literature on face recognition have been reviewed. From the literature review, as documented in section 2.2, it is obvious that vehicle MMR is a relatively unexplored research area that can benefit from further research. The feature extraction techniques can be further improved to include features which better describe the geometrical structure and contours along the curves of an image of a typical car. Further, local and global description of the car images need to be enhanced. Instead of using all features in classification, the selection of the most representative features can increase recognition rates and reduce the inaccuracies that may result from noise. Further matching criteria used in the current literature on vehicle MMR can be improved for example by replacing simple keypoint to keypoint matching in SIFT based methods, with a more reliable matching scheme which identifies similar feature points within maximum likelihood areas.
Considering the above shortcomings of the existing algorithms, a number of novel approaches to vehicle MMR have been proposed in chapters 4-7. The research motivation behind each approach has been presented in the relevant chapters. Chapter 3 introduces the reader to the fundamental concepts/theories used within the above contributory chapters.
CHAPTER 3

Research Background

3.0 INTRODUCTION

The vehicle recognition algorithms proposed in this thesis aim to overcome the limitations of existing vehicle recognition techniques, in order to perform effectively with better accuracy. This chapter introduces the reader to the fundamental concepts/theories based on which the proposed algorithms have been developed. Further section 3.12 provides details of the car image database that was used in experiments.

3.1 STATISTICAL PRELIMINARIES

Covariance: Covariance $\text{cov}(X,Y)$ of two (random) variables is defined as to be:

$$\text{cov}(X,Y) = \frac{\sum_{i=1}^{n} (X_i - \bar{X})(Y_i - \bar{Y})}{n-1}$$  \hspace{1cm} (3.1)

Where $\bar{X}$ and $\bar{Y}$ represent the mean of variables $X$ and $Y$ respectively.

Since the covariance value can be calculated between any 2 dimensions in a data set, this technique is often used to find relationships between dimensions in high-dimensional data sets where visualisation is difficult.
Covariance Matrices: In general, for an \( n \)-dimensional data set, one can calculate \( \binom{n}{2} = \frac{n(n-1)}{2} \) different covariance values.

A useful way to obtain all the possible covariance values between all the different dimensions \( D_{\text{Dim}} \) is to calculate them all and place them in a matrix. Definition for the covariance matrix for a set of data with \( n \) dimensions is:

\[
C^{n \times n} = (C_{i,j}, C_{i,j} = \text{cov}(\text{Dim}_i, \text{Dim}_j))
\]

(3.2)

Where \( C^{n \times n} \) is a matrix with \( n \) rows and \( n \) columns. This formula indicates that given a \( n \) dimensional data set, the covariance matrix has \( n \) rows and \( n \) columns and each entry in the matrix is the covariance between two separate dimensions. For example entry on row 2, column 3, is the covariance value calculated between the 2nd dimension and the 3rd dimension.

As an example, consider the covariance matrix for an imaginary 3 dimensional data set, using the usual dimensions \( x, y \) and \( z \). Then, the covariance matrix has 3 rows and 3 columns, and the values are:

\[
\begin{bmatrix}
\text{cov}(x,x) & \text{cov}(x,y) & \text{cov}(x,z) \\
\text{cov}(y,x) & \text{cov}(y,y) & \text{cov}(y,z) \\
\text{cov}(z,x) & \text{cov}(z,y) & \text{cov}(z,z)
\end{bmatrix}
\]

(3.3)

It is observed that

- Down the main diagonal, covariance value is between one of the dimensions and itself. These are the variances for that dimension.
- \( \text{cov}(x, y) = \text{cov}(y, x) \) the matrix is symmetrical about the main diagonal.
Eigenvectors and Eigenvalues: Definition: Let A be a complex square matrix. Then if \( \lambda \) is a complex number and \( u \) a non-zero complex column vector satisfying \( Au = \lambda u \), \( u \) is called an eigenvector of \( A \), while \( \lambda \) is called an eigenvalue of \( A \). Further \( u \) is said to be an eigenvector corresponding to the eigenvalue \( \lambda \). Details on how to find eigenvectors can be found in (Anton)[47].

3.2 PRINCIPAL COMPONENT ANALYSIS (PCA)

Principal Components Analysis (PCA) was introduced in 1901 by Karl Pearson[48]. It is a way of identifying patterns in data or images, and expressing the data (including images) in such a way as to highlight similarities and differences. The main idea is to reduce the dimensionality of a data set which consists of a large number of inter-related variables, while retaining as much as possible of the variation present in the original data set. Since patterns in data can be hard to find in data of high dimension - PCA is a powerful tool for analysing data.

The reduction in dimensionality is achieved by transforming to a new set of variables, the principal components (PCs), which are uncorrelated (i.e. they are linearly independent), and which are ordered so that the first few retain most of the variation present in all of the original variables. This technique is used in image compression, pattern recognition, data mining and other applications.

3.2.1 PCA for Two Variables

Though PCA is normally applied to more than two variables, its use is demonstrated using two variables, denoted by \( X \) and \( Y \). The purpose of PCA is to find axis, or directions, in a data set that are the most significant with regard to representing the data. Such an axis is illustrated with help of direction in Figure 3.1 and is identified as being that along which there is maximum variance (or scatter) - compare with Figure 3.2.
Figure 3.3 shows the ‘natural’, or principal, axis system for the data set \((X,Y)\) relative to the original axis system. Figure 3.4 illustrates the data set relative to its principal axis system - origin of the new system is at \((\bar{X},\bar{Y})\).
The ideas and nomenclature come from geometry; Figure 3.5 shows an ellipse in an arbitrary axis system and in its principal axis system. The origin of the principal axis system is the centre of gravity of the ellipse - \((\bar{x}, \bar{y})\) is the analogue of the centre of gravity for data sets. The principal axis system is the natural system for the ellipse in the sense that the ellipse is symmetrical in the principal system. The principal axis system for a 2- or 3-dimensional object may be found by diagonalising its inertia tensor by means of eigenanalysis; the principal system for a data set is found by diagonalising its covariance tensor (matrix) also using eigenanalysis - see section 3.2.2.
The variable measured along the axis of maximum variance axis is a linear combination, $\alpha X + \beta Y$ of the original variables - called the first principal component. Therefore the need is to determine suitable constants $\alpha$ and $\beta$ such that $\text{var}(\alpha X + \beta Y)$ is maximized.

### 3.2.2 Computing Principal Components of PCA

Given that

$$\text{var}(\alpha X + \beta Y) = \alpha^2 \text{var}(X) + \beta^2 \text{var}(Y) + 2\text{cov}(X,Y)$$

(3.4)

Principal components $P_1$ and $P_2$ are computed as follows:

- Find linear combination $P_1 = \alpha_1 X + \beta_1 Y$, where $\| (\alpha_1, \beta_1) \| = 1$, that has maximum variance.
CHAPTER 3

- Find a linear combination $P_2 = \alpha_2 X + \beta_2 Y$, uncorrelated with $P_1$ (i.e. $\text{cov}(P_1, P_2) = 0$), where $\|\alpha_2, \beta_2\|=1$, which has maximum variance.
- For two variables no further uncorrelated linear combinations will exist; $P_1$ is called the first principal component of $(X, Y)$ and $P_2$ is the second principal component.

3.2.3 The Use of PCA in Feature Extraction and Classification

This section explains the stepwise procedure of computing PCA. The feature extraction from images is a complex problem as the data are often high dimensional and it is hard to find patterns. Therefore the need is to first transform it to a lower dimensional space and then the recognition.

PCA computes the basis of a space which is represented by its training vectors. These basis vectors, actually eigenvectors, computed by PCA are in the direction of the largest variance of the training vectors. Each eigenvector can be viewed as a feature. When an image of an object is projected onto the eigenspace, its feature vector projections describe the importance of each of those features in the image. The image is expressed in the eigenspace by its eigenvector coefficients (or weights). For a large input vector; an image can be handled only by taking its small weight vector in the eigenspace. This means that the original image can be reconstructed with some error, since the dimensionality of the image space is much larger than that of eigenspace [49]. In practice PCA provides a fast, model free approach to feature extraction and classification.

According to [49] a step-by-step approach to PCA based feature extraction and classification can be presented as follows:
Step 1: Obtain \( E \) images \((N \times N \) dimension each) represented by \( I_1, I_2, \ldots, I_M \) for training.

Step 2: Convert each Image \( I_i \) to \( \Gamma_i \) \((N^2 \times 1) \) vector. As a result \( \Gamma_1, \Gamma_2, \ldots, \Gamma_E \) vectors are obtained.

Step 3: compute the mean/average image vector \( \Psi \) as

\[
\Psi = \frac{1}{E} \sum_{i=1}^{E} \Gamma_i
\]  

(3.5)

Step 4: obtain mean subtracted image vectors by subtracting the mean vector \( \Psi \) from each vector \( \Gamma_i \) and call it \( \Phi_i \), i.e.

\[
\Phi_i = \Gamma_i - \Psi
\]  

(3.6)

Step 5: Calculate the covariance matrix

\[
C = \frac{1}{E} \sum_{i=1}^{E} \Phi_i \Phi_i^T = BB^T \quad (N^2 \times N^2 \text{ matrix})
\]  

(3.7)

Where

\[
B = [\Phi_1 \Phi_2 \ldots \Phi_E] \quad (N^2 \times E \text{ matrix})
\]  

(3.8)

Step 6: calculate the eigenvectors \( u_i \) for covariance matrix \( C = BB^T \). Note that \( BB^T \) is a very large matrix. Consider \( B^T B \) \((E \times E \text{ matrix})\). Unit Eigenvectors \( v_i \) of \( B^T B \) are computed instead.

\[
B^T B v_i = \lambda_i v_i
\]  

(3.9)

Relationship between \( u_i \) and \( v_i \) is given by:
CHAPTER 3

\[ B^T B v_i = \lambda_i v_i \Rightarrow B B^T B v_i = \lambda_i B v_i \Rightarrow C B v_i = \lambda_i B v_i, \]

or

\[ C u_i = \lambda_i u_i, \]

where

\[ u_i = B v_i, \]  \hspace{1cm} (3.10)

So \( B^T B \) and \( B B^T \) has same eigenvectors and eigenvalues and related as \( u_i = B v_i \)

**Note:**

\( B B^T \) can have up to \( N^2 \) eigenvectors and eigenvalues
\( B^T B \) can have up to \( E \) eigenvectors and eigenvalues

Compute \( E \) most significant largest value eigenvectors and eigenvalues for \( B B^T \).

**Step 7:** select and keep only \( K \) eigenvectors corresponding to \( K \) largest eigenvalues.

**Step 8:** Project images on this eigenspace: Each image (mean subtracted) \( \Phi_i \) in the training set can be represented as the linear combination of \( K \) eigenvectors.

\[ \hat{\Phi}_i = \sum_{j=1}^{K} w_j u_j, \quad (w_j = u_j^T \Phi_i) \]  \hspace{1cm} (3.11)

Each training image \( \Phi_i \) is represented by a vector in eigenspace

\[ \Omega_i = \begin{bmatrix} w'_1 \\ . \\ . \\ w'_K \end{bmatrix}, \quad i = 1,2,...,E \]  \hspace{1cm} (3.12)

**Step 9:** Take an unknown test image represented as \( \Gamma \) to match with the database of training images. \( \Gamma \) has been obtained the same way as in step2 and normalised by subtracting the mean image (from step3) of training set.
Step 10: Project it on the eigenspace

$$\hat{\Phi}_i = \sum_{i=1}^{K} w_i u_i , \quad (w_i = u_i^T \Phi)$$

(3.13)

Step 11: Represent $\Phi$ as

$$\Omega = \begin{bmatrix} w_1 \\ \cdot \\ \cdot \\ w_K \end{bmatrix}$$

Step 12: Compute the Euclidean distance as the matching error between the test image and each of the training images, using:

$$err = \min_{j} \| \Omega - \Omega_j \|$$

(3.14)

If the error, ‘err’ becomes less than some threshold ‘$T$’, test image matches best with training image ‘$i$’ and is given the class label of that training image.

3.3 LINEAR DISCRIMANT ANALYSIS (LDA)

In section 3.2 it was discussed that PCA is one of the most frequently used approaches for dimensionality reduction. PCA searches for directions in the data that have largest variance and subsequently project the data onto those directions. In this way, a lower dimensional representation of the data is obtained, that removes some of the ‘noisy’ directions. However, PCA does not include class/label information of the data. For instance, if we imagine 2 cigar like clusters in 2 dimensions, one cigar has class label ‘1’ and the other has class label ‘-1’. The cigars are positioned in parallel and very closely together, such that the variance in the total data-set, ignoring the labels, is in the direction of the cigars. For classification, this would be an inappropriate projection, because all labels become
evenly mixed and useful information is destroyed. A much more useful projection is orthogonal to the cigars, i.e. in the direction of least overall variance, which would perfectly separate the data-cases (classification in this 1-D space is still needed).

Apart from PCA, there are a number of further techniques that can be used for the classification of data. Linear Discriminant Analysis (LDA) is one of the most commonly used techniques. LDA was originally developed in 1936 by R.A. Fisher [50]. This method is used in statistics and machine learning to find the linear combination of features which best separate two or more classes of objects or events. The resulting combination may be used as a linear classifier or, more commonly, for dimensionality reduction before later classification. LDA explicitly attempts to model the difference between the classes of data. Generally LDA is preferred over PCA from the classification point of view because; the former deals directly with discrimination between classes, whereas the latter deals with the data in its entirety for the principal components analysis without paying any particular attention to the underlying class structure. The prime difference between PCA and LDA is that the former performs feature classification and later performs data classification. In PCA, the shape and location of the original data sets changes when transformed to a different space whereas LDA does not change the location but only tries to provide more class separability and draws a decision region between the given classes [51]. The comparison of PCA and LDA can be visualized as in Figure 3.6.
3.3.1 Theory of LDA

In LDA, similar to in PCA, the image is initially linearly projected into a feature subspace. The projection method is based upon Fisher’s Linear Discriminant Analysis and produces well separated classes in a low-dimensional subspace. The objective of LDA is to find the optimal projection so that the ratio of the determinants of the between-class and the within-class scatter matrices of the projected samples reaches its maximum [50]. It also attempts to effectively draw a decision region between the given classes. Mathematically speaking, for all images of all classes, two measures are defined according to [52]:

1) Within-class scatter matrix showing the average scatter of samples images $I_s$ of different classes $H_i$ around their respective means $\Psi_i$, as given by:
\[
S_w = \sum_{i=1}^{h} \sum_{k \in H_i} (\Psi_k - \Psi)(\Psi_k - \Psi)^T
\]  
(3.15)

Here \( h \) represents the number of classes.

2) Between-class scatter matrix representing the scatter of the individual means \( \Psi_i \) around overall mean \( \Psi \), as given by:

\[
S_B = \sum_{i=1}^{h} Z_i (\Psi_i - \Psi)(\Psi_i - \Psi)^T
\]  
(3.16)

Where \( Z_i \) represents the number of samples in class \( H_i \). If \( S_w \) is non-singular, the optimal projection \( W_{opt} \) is chosen as the matrix with orthonormal columns which maximises the ratio of the determinant of between-class scatter matrix of the projected samples to the determinant of the within-class scatter matrix of the projected samples, i.e.

\[
W_{opt} = \arg \max_W \frac{|W^T S_B W|}{|W^T S_w W|}
\]

\[
= [u_1, u_2, ..., u_m]
\]  
(3.17)

Where \( \{u_i | i = 1, 2, ..., m\} \) is the set of generalised eigenvectors of \( S_B \) and \( S_w \) corresponding to the \( m \) largest generalised eigenvalues \( \{\lambda_i | i = 1, 2, ..., m\} \)

\[
S_B u_i = \lambda_i S_w u_i, \quad i = 1, 2, ..., m
\]  
(3.18)

There are at most \( h - 1 \) nonzero generalised eigenvectors.
3.4 TWO DIMENSIONAL LINEAR DISCRIMANT ANALYSIS (2DLDA)

Linear Discriminant Analysis (LDA) presented in section 3.3 has been used widely in many applications involving high-dimensional data, such as image retrieval [53] and face recognition [54]. An intrinsic limitation of classical LDA is the singularity problem, that is, it fails when all scatter matrices are singular. A well-known approach to deal with the singularity problem is to apply an intermediate dimension reduction stage using PCA before LDA. The algorithm, known as PCA+LDA, is used widely in face recognition [54]. However, PCA+LDA has high costs in time and space, due to the need for an eigen-decomposition involving the scatter matrices. To overcome this problem, a novel LDA extension algorithm, namely 2DLDA, which stands for Two Dimensional Linear Discriminant Analysis has been proposed by M.Li and B.Yuan [55]. 2DLDA overcomes the singularity problem implicitly, while achieving efficiency. The key difference between 2DLDA and classical LDA lies in the model for data representation. Classical LDA works with vectorized representations of data, while the 2DLDA algorithm works with data in matrix representation. Matrix representation in 2DLDA leads to an eigen-decomposition on matrices which are much smaller than the matrices in classical LDA. This significantly reduces the time and space complexities of 2DLDA over LDA [56].

3.4.1 2DLDA Algorithm for Feature Extraction

According to [55], a step-by-step approach to 2DLDA based feature extraction and classification can be presented as follows:

Step 1: Suppose there are \( h \) known pattern classes in the training set of images, and \( E \) denotes the size of the training set. Denote \( j^{th} \) training image by a \( m \times n \) matrix \( I_j (j = 1,2,\ldots, E) \).
Step 2: Compute the mean image for each class $H_i$ and denote it by $\Psi_i (i = 1, 2, \ldots, h)$. Consider $Z_i$ as the number of samples in class $H_i$.

Step 3: Compute the mean image of all training samples and denote it by $\Psi$.

Step 4: Let $u$ denotes an n-dimensional column vector. The training image $I_j$ is projected onto $u$ and a projected feature vector $y_j$ is obtained by the following linear transformation.

$$y_j = I_j u \quad \quad j = 1, 2, \ldots, E$$  \hspace{1cm} (3.19)

According to [49], the total scatter of the projected samples can be characterized by the trace [57] of the covariance matrix of the projected feature vectors.

$$J(u) = \frac{tr(TS_B)}{tr(TS_w)}$$  \hspace{1cm} (3.20)

where $TS_B$ denotes the between-class scatter matrix of projected feature vectors of training images, and $TS_w$ denotes the within-class scatter matrix of projected feature vectors of training images.

Step 5: compute $TS_B$ and $TS_w$ as follows:

$$TS_B = \sum_{i=1}^{h} Z_i (\bar{y}_i - \bar{y})(\bar{y}_i - \bar{y})^T$$  \hspace{1cm} (3.21)

Where $\bar{y}_i$ represents projected feature vector of mean image of class $H_i$, and $\bar{y}$ represents the projected feature vector of mean image of all training samples.

Substituting $\bar{y}_i = \Psi_i u$ and $\bar{y} = \Psi u$ in (3.21)
TS_B = \sum_{i=1}^{h} Z_i [(\Psi_i - \Psi)u] [(\Psi_i - \Psi)u]^T

TS_W = \sum_{i=1}^{h} \sum_{y_k \in H_i} (y_k - \bar{y}_i)(y_k - \bar{y}_i)^T

(3.22)

Where \( y_k \) represents projected feature vector of a training image of class \( H_i \).

Substituting \( y_k = I_k u \) and \( \bar{y}_i = \Psi_i u \) in (3.22)

\[
TS_W = \sum_{i=1}^{h} \sum_{I_k \in H_i} [(I_k - \Psi_i)u][(I_k - \Psi_i)u]^T
\]

Trace of both matrices will be

\[
tr(TS_B) = u^T \sum_{i=1}^{h} Z_i (\Psi_i - \Psi)^T (\Psi_i - \Psi)u
\]

\[
= u^T S_B u
\]

(3.23)

\[
tr(TS_W) = u^T \sum_{i=1}^{h} \sum_{I_k \in H_i} (I_k - \Psi_i)^T (I_k - \Psi_i)u
\]

\[
= u^T S_W u
\]

(3.24)

(3.20) can now be expressed as

\[
J_{(u)} = \frac{u^T S_B u}{u^T S_W u}
\]

(3.25)

Where \( u \) is a unitary column vector. This criterion is called Fisher linear projection criterion. The unitary vector \( u \) that maximizes \( J_{(u)} \) is called the Fisher optimal projection axis. The optimal projection \( u_{opt} \) is chosen when the criterion is maximized, i.e.

\[
u_{opt} = \arg \max_u J_{(u)}
\]

(3.26)
Step 6: If $S_w$ in non-singular, compute solution to (3.26) using generalised
eigenvalue problem as:

$$S_w^{-1}S_B u_{opt} = \lambda u_{opt} \quad (u_{opt} \text{ is optimal Fisher projection axis}) \quad (3.27)$$

In general, set of projection axis $\{u_1, u_2, ..., u_r\}$ is required. These projection axes are
the orthonormal eigenvectors of $S_w^{-1}S_B$. [Note $r$ represents the number of
eigenvectors with largest eigenvalues].

Step 7: Place all the eigenvectors in a matrix to obtain the fisher projection matrix
$n \times r$ as,

$$U = [u_1, u_2, ..., u_r] \quad (3.28)$$

Step 8: Extract features for an image $I$ by obtaining a feature matrix $Y$ of
dimension $m \times r$, using the projection Vectors $\{u_1, u_2, ..., u_r\}$ derived in step 6 as follows:

$$Y = [y_1, y_2, ..., y_r] \quad (3.29)$$

Where

$$y_k = Iu_k \quad k = 1, 2, ..., r$$

Using feature matrix and fisher optimal projection axis, an image can be
reconstructed. Further details can be found in [55].

3.4.2 2DLDA and Singularity

2DLDA successfully overcomes the singularity issue found in traditional LDA. The
reason is that for each training image $I_j (j = 1, 2, ..., E)$, the rank is defined as:
CHAPTER 3

\[ \text{rank}(I_J) = \min(m, n) \]  
(3.30)

From (3.24)

\[ \text{rank}(S_w) = \text{rank}\left(\sum_{i=1}^{k} \sum_{i \neq i'} (I_K - \Psi_i)^T (I_K - \Psi_{i'})\right) \leq (E - h).\min(m, n) \]  
(3.31)

\( S_w \)  Non-singular when

\[ E \geq h + \frac{n}{\min(m, n)} \]  
(3.32)

According to M. Li and B. Yuan [55], in a real situation (3.32) is always satisfied. Therefore in 2DLDA, \( S_w \) is non-singular.

3.5 SCALE INVARIANT FEATURE TRANSFORM (SIFT)

Scale Invariant Feature Transforms, first proposed by David Lowe [58] is an approach that has the ability to extract distinctive, invariant features from images. These features can be used for reliable matching of images, thereby making SIFT highly suitable for robust image retrieval and object recognition applications [59, 60]. It has been shown that SIFT features are theoretically fully-invariant to scale and rotation and that they are partially invariant (i.e. robust) to affine distortion, noise, change in 3D viewpoint and illumination [58, 59]. A significant number of features can be extracted from typical images, which densely cover the image over the full range of scales and locations. For example, a typical image of size 500*500 pixels may give rise to about 2000 stable features (the exact number depends on both image content and choices of parameters). Further the number of feature points selected can be flexibly varied, by the suitable selection of parameter values. In addition, the SIFT features are highly distinctive, which allows a single feature to be correctly matched with high probability against a large database of features, providing a solid basis for object identification/recognition [60].
Chapter 3

The key stages of SIFT can be briefly summarized as follows [58]:

3.5.1 Detection of Scale-space Extrema

Scale-space construction: The first stage of computation involves the search for stable features/keypoints over all scales and image locations. The scale space of an image is defined as a function $L(x, y, \sigma)$ which is produced from the convolution of a variable-scale Gaussian, $G(x, y, \sigma)$ with an input image $I(x, y)$:

$$L(x, y, \sigma) = G(x, y, \sigma) * I(x, y)$$  (3.33)

Where $*$ is a convolution operator

$$G(x, y, \sigma) = \frac{1}{2\pi\sigma^2} e^{-\frac{(x^2+y^2)}{2\sigma^2}}$$  (3.34)

Stable keypoint locations in scale space are found using extrema in the Difference-of-Gaussian (DoG) function convolved with the image, $Diff(x, y, \sigma)$, which can be computed from the difference of two nearby scales (filtered images) separated by a constant multiplicative factor $k$:

$$Diff(x, y, \sigma) = L(x, y, k\sigma) - L(x, y, \sigma)$$  (3.35)

Figure 3.7 illustrates the formation of a scale space for a given image. For each octave of scale space, the initial image is repeatedly convolved with Gaussian filters to produce the set of scale space images shown on the left side of Figure 3.7. Adjacent Gaussian images are subtracted to produce the DoG images shown on the
right side of Figure 3.7. After each octave, the Gaussian image with $\sigma$ twice that of the original is sub-sampled and used to construct the next octave.

![Diagram of Gaussian blurred images and their respective DoG images](image)

Figure 3.7: Gaussian blurred images at different scales and their respective DoG images [58].

Details on selection of number of images within each octave of scale-level and value of $k$ can be found in [58].

**Extrema Detection:** An extrema is defined as any value in the DoG greater than or smaller than all its neighbours in scale-space. To find the extrema points each pixel is compared to 8 neighbours at the same scale plus 9 neighbours each from scales above and below (see Figure 3.8). If the pixel is a local maximum or minimum, it is selected as a candidate keypoint.
3.5.2 Keypoint Localisation

The candidate keypoints selected using the procedure outlined in section 3.5.1 are further refined (keypoint localization) by measuring their stability. A detailed model is subsequently fit to pixels surrounding the candidate keypoints, for determining location, scale and principal of curvature. By using this information, keypoints having low contrast or are poorly localized along edges are rejected. The remaining keypoints are classified as stable keypoints and are used in subsequent analysis [58].

3.5.3 Orientation Assignment

Each of the stable keypoints is then assigned with an orientation. The orientation of a keypoint is calculated by computing a gradient orientation histogram for pixels in the neighbourhood of the keypoint [58] as illustrated in Figure 3.9. The peaks in the histogram are considered to correspond to the dominant orientations. The maximum value and the 80% of the maximum value in histogram are used to create separate keypoints for these directions.
CHAPTER 3

3.5.4 Keypoint Descriptor Computation

The orientation of the keypoint is then used to find out the feature descriptor of the keypoint. In this stage, a descriptor is computed for the local image region that is as distinctive as possible at each keypoint. The image gradient magnitudes and orientations are sampled around the keypoint location. These values are illustrated with small arrows at each sample location on the left-hand image in Figure 3.10. The contribution of each pixel is weighted by the gradient magnitude, and by a Gaussian weighting function with $\sigma$ 1.5 times the scale of the keypoint. Each descriptor is a 128 ($8 \times 4 \times 4$) element feature vector [58].

Figure 3.10: Left: the gradient magnitude and orientation at a sample point in a square region around the keypoint location. These are weighted by a Gaussian window, indicated by the overlaid circle. Right: The image gradients are added to an orientation histogram. Each histogram include 8 directions indicated by the arrows and is computed from 4x4 sub-regions. The length of each arrow corresponds to the sum of the gradient magnitudes near that direction within the region [58].
Details of the stages of SIFT can be found in [58] and [59].

3.5.5 Invariance Properties of SIFT

Scale Invariance is achieved for SIFT keypoints by extracting and selecting the point of extrema which remain stable across the whole scale-space of DoG images. In order to achieve rotation invariance for the keypoint, the coordinates of descriptor and gradient orientations are rotated relative to that of the keypoint orientation. To make the keypoint illumination invariant, the feature vector for the descriptor is modified. The illumination variations could be of several types and therefore the feature vector is modified accordingly. For example a change in image contrast in which each pixel value is multiplied by a constant, will multiply gradients by the same constant, so this contrast change will be cancelled by vector normalization to unit length. A brightness change in which a constant is added to each image pixel will not affect the gradient values, as they are computed from pixel differences. Therefore, the descriptor is invariant to affine changes in illumination. However, non-linear illumination changes can also occur due to camera saturation or due to illumination changes that affect 3D surfaces with differing orientations by different amounts. These effects can cause a large change in relative magnitudes for some gradients, but are less likely to affect the gradient orientations. Therefore, influence of large gradient magnitudes is reduced by thresholding the values in the unit feature vector, and then renormalizing to unit length [58].

3.6 CURVELET TRANSFORMS

Curvelets as proposed by E. Cand Cand’es and D. Donoho [61], constitute a relatively new family of frames (i.e. sub-bands) that are designed to represent edges and other singularities along signed curves much more efficiently than the traditional wavelet based transforms. The curvelet transform is a multi-scale
transform with frame elements indexed by location, scale and orientation parameters, and have time-frequency localization properties of wavelets but also shows a very high degree of directionality and anisotropy. The Continuous Curvelet Transform has gone through two major revisions.

1st Generation curvelet transforms: The first Continuous Curvelet Transform [61] (commonly referred to as the 'Curvelet 99' transform now) used a complex series of steps involving the ridgelet [62, 63] analysis of the radon transform of an image.

The procedural definition of Curvelet Transform [61, 64] involving ridgelet analysis step includes four stages:

3.6.1 Stages of Curvelet Transform

Sub-band decomposition: For a bank of sub-band filters $P_0, \Delta_s (s \geq 0)$, an object/image $f$ is divided into resolution layers whereas each layer contains details of different frequencies:

$$f \rightarrow (P_0 f, \Delta_1 f, \Delta_2 f, \ldots)$$

$P_0$ - Low-pass filter

$\Delta_1, \Delta_2, \ldots$ - Band-pass (high-pass) filters

The original image can be reconstructed from the sub-bands:

$$f = P_0 (P_0 f) + \sum_s \Delta_s (\Delta_s f)$$

(3.36)

The different sub-bands $\Delta_s f$ contain details about $2^{-2s}$ wide.

Smooth partitioning: A collection of smooth windows $w_q(x_1, x_2)$ localized around dyadic squares
CHAPTER 3

\[ Q(s, k_1, k_2) = \left[ \frac{k_1}{2^s}, \frac{k_1 + 1}{2^s} \right] \times \left[ \frac{k_2}{2^s}, \frac{k_2 + 1}{2^s} \right] \in Q_s \quad Q_s - \text{all the dyadic squares of the grid} \]

are defined. Multiplying a function by the corresponding window function \( w_Q \) produces a result localized near \( Q \). Doing this for all \( Q \) at a certain scale, i.e. for all \( Q = Q(s, k_1, k_2) \) with \( k_1 \) and \( k_2 \) varying but \( s \) fixed, produces a smooth dissection of the function into 'squares'. In this stage of the procedure, this windowing dissection is applied to each of the sub-bands isolated in the previous stage of the algorithm (section 3.6.1). 'Squares' of appropriate scales are obtained.

\[ \Delta_s f \rightarrow (w_Q \Delta_s f)_{Q \in Q_s} \]

Renormalization: Each resulting dyadic square is centered to the unit square \([0, 1] \times [0, 1]\). For each \( Q \), the operator \( T_Q \) which transports and renormalizes \( f \) is defined as:

\[ (T_Q f)(x_1, x_2) = 2^s f(2^s x_1 - k_1, 2^s x_2 - k_2) \]  \hspace{1cm} (3.37)

Each square is renormalized:

\[ g_Q = 2^{-s}(T_Q)^{-1}(w_Q \Delta_s f) \quad Q \in Q_s \]  \hspace{1cm} (3.38)

Ridgelet analysis: Each 'square' is analyzed in the orthonormal ridgelet system [62, 63]. This is a system of basis elements \( \rho_\lambda \), making an orthobasis for \( L^2(\mathbb{R}^2) \):

\[ \sigma_\mu = (g_Q, \rho_\lambda), \mu = (Q, \lambda) \]  \hspace{1cm} (3.39)

The flow of these stages and an example is illustrated in the Figure 3.11 and Figure 3.12 respectively.
CHAPTER 3

Figure 3.11: Organization of curvelet transforms [64]

Figure 3.12: An example Spatial decomposition of single sub-band [64].
2nd Generation curvelet transforms: The Performance of 'Curvelet 99' was slow. The algorithm was updated in 2002 by E.Cand'es and D. Donoho [65]. The use of the Ridgelet Transform was discarded, thus reducing the amount of redundancy in the transform and increasing the speed considerably. In this new method, an approach of curvelets as tight frames is taken. Using tight frames, an individual curvelet has frequency support in a parabolic-wedge area of the frequency domain (As seen in Figure 3.13.).

![Figure 3.13: Curvelet tiling of the frequency plane. In frequency domain curvelets are supported near symmetric 'parabolic' wedges. The shaded area represents this wedge [65]](image-url)

Using the theoretical basis in [65] (where the continuous curvelet transform is created), two separated digital (or discrete) curvelet transform (DCT) algorithms are introduced in [66]. The first algorithm is the Unequispaced Fast Fourier Transform (FFT) [65], where the curvelet coefficients are found by irregularly sampling the fourier coefficients of an image. The second algorithm is the Wrapping Transform [65], using a series of translations and a wraparound technique. Both algorithms have the same output, but the Wrapping Algorithm gives both, a more intuitive algorithm and faster computation time.
3.6.2 Why Use Curvelets?

The curvelet transform, like the wavelet transform, is a multi-scale transform, with frame elements indexed by scale and location parameters. Unlike the wavelet transform, it has directional parameters, and the curvelet pyramid contains elements with a very high degree of directional specificity. In addition, the curvelet transform is based on a certain anisotropic scaling principle which is quite different from the isotropic scaling of wavelets. The elements obey a special scaling law, where the length of the support of a frame elements and the width of the support are linked by the relation width $\propto$ length$^2$. Wavelets do well for point singularities, and not for singularities along curves. The success of wavelets in dimension 1 is derived from the fact that all singularities in dimension 1 are point singularities, so wavelets have certain universality. In higher dimensions there are more types of singularities, and wavelets lose their universality. The Curvelet transform was designed to represent edges and other singularities along curves much more efficiently than traditional transforms, i.e. using many fewer coefficients for a given accuracy of reconstruction. In general, to represent an edge to squared error $\sqrt{N}$ requires $\sqrt{N}$ wavelets and only about $\sqrt{\sqrt{N}}$ curvelets [64].

When the curvelet transform is performed on a $C^2$ curve, a very few curvelet coefficients will be above negligible magnitude values. In [65], it is declared that curvelets offer optimal sparseness for 'curve-punctuated smooth' images, where the image is smooth with the exception of discontinuities along $C^2$ curves. Sparseness is measured by the rate of decay of the m-term approximation (reconstruction of the image using $m$ number of coefficients) of the algorithm. According to D.L.Donoho [67] sparse representation along with offering improved compression possibilities, also allows for improving denoising performance, as additional sparseness increases the amount of smooth areas in the image. In [68] it was shown that orthogonal systems have optimal m-term approximations that decay in $L^2$ with rate $O(m^{-2})$ (as a lower bound). Currently, a single computationally feasible transform that will
CHAPTER 3

obtain this lower bound does not exist. On images with $C^2$ boundaries, non-optimal systems have the following rates:

Fourier Approximation

$$\|f - f_m^F\|_{L^2} \approx O(\frac{1}{m^2})$$  \hspace{1cm} (3.40)

Wavelet Approximation

$$\|f - f_m^W\|_{L^2} \approx O(m^{-1})$$  \hspace{1cm} (3.41)

Curvelet Approximation

$$\|f - f_m^C\|_{L^2} \approx O((\log m)^{3} m^{-2})$$  \hspace{1cm} (3.42)

As seen from the m-term approximations, the Curvelet Transform offers the closest m-term approximation to the lower bound. Therefore, in images with a large number of $C^2$ curves (i.e. an image with a great number of long edges), it would be advantageous to use the Curvelet Algorithm.

3.6.3 Challenges in Using Curvelets

Curvelet constructions [61, 65] require a rotation operation and correspond to a 2-D frequency partition based on the polar coordinate. This makes the curvelet construction simple in the continuous domain but causes the implementation for discrete images (sampled on a rectangular grid) to be very challenging. In particular, approaching critical sampling seems difficult in such discretized constructions.

3.7 CONTOURLET TRANSFORMS

The main challenge in exploring geometry in images comes from the discrete nature of the data. Thus, unlike other approaches, such as curvelets, that first develop a
transform in the continuous domain and then discretize for sampled data, contourlet transform by M. N. Do and M. Vetterli [69] starts with a discrete-domain construction and then studies its convergence to an expansion in the continuous domain. Specifically, a discrete-domain multi-resolution and multi-direction expansion using non-separable filter banks is constructed, in much the same way that wavelets are derived from filter banks. This construction results in a flexible multi-resolution, local, and directional image expansion using contour segments, and, thus, it is named the 'contourlet' transform. Performance of wavelet and contourlet transform near smooth contours is illustrated below in Figure 3.14.

![Figure 3.14: Wavelet versus Contourlet illustrating the successive refinement by the two systems near a smooth contour, this is shown as a thick curve separating two smooth regions [69]](image)

### 3.7.1 Discrete Domain Construction

Comparing the wavelet scheme with Contourlet scheme (see Fig. 3.14), it is observed that the improvement of the contourlet scheme can be attributed to the grouping of nearby wavelet coefficients, since they are locally correlated due to the smoothness of the contours. Therefore, a sparse expansion for natural images is obtained by first applying a multi-scale transform, followed by a local directional transform to gather the nearby basis functions at the same scale into linear structures. In this regard, a 'double filter bank' structure for obtaining sparse expansions for typical images having smooth contours has been proposed [70]. The stages involved in double filter bank are:
Multi-scale decomposition: In this stage, the Laplacian Pyramid (LP) introduced by Burt and Adelson [71] is used. The LP decomposition at each level generates a down sampled lowpass version of the original and the difference between the original and the prediction, resulting in a band-pass image. Fig. 3.15 depicts this decomposition process.

The LP has the distinguishing feature that each pyramid level generates only a one band-pass image (even for multidimensional cases), and this image does not have 'scrambled' frequencies. This frequency scrambling happens in the wavelet filter bank when a highpass channel, after downsampling, is folded back into the low frequency band, and, thus, its spectrum is reflected. In the LP, this effect is avoided by downsampling the lowpass channel only.

Directional Filter Bank (DFB): DFB in M. N. Do [72] is constructed from two building blocks. The first building block is a two-channel quincunx filter bank [73] fan filters (see Figure. 3.17) that divides a 2-D spectrum into two directions: horizontal and vertical. The second building block of the DFB is a shearing operator, which amounts to just reordering of image samples. Thus, the key in the DFB is to use an appropriate combination of shearing operators together with two-direction partition of quincunx filter banks at each node in a binary tree-structured filter bank, to obtain the desired 2-D spectrum division as shown in Figure. 3.16. Mathematical details can be found in [69].

Figure 3.15. LP: One level of decomposition. The outputs are a coarse approximation $c[n]$ and a difference $d[n]$ between the original signal and the prediction. $H$ and $G$ are called (lowpass) analysis and synthesis filters respectively and $M$ is a sampling matrix [69].
Combining LP and DFB: LP and DFB obtained are combined into a double filter bank structure. Since the DFB was designed to capture the high frequency (representing directionality) of the input image, the low frequency content is poorly handled. In fact, with the frequency partition shown in Figure 3.16, low frequency would 'leak' into several directional sub-bands; hence, the DFB alone does not provide a sparse representation for images. Figure 3.18 shows a multi-scale and directional decomposition using a combination of a LP and a DFB. Band-pass images from the LP are fed into a DFB so that directional information can be captured. The scheme can be iterated on the coarse image. The combined result is a
double iterated filter bank structure, named contourlet filter bank, which decomposes images into directional sub-bands at multiple scales [69].

![Contourlet filter bank diagram](image)

**Figure 3.18**: Contourlet filter bank First, a multi-scale decomposition into octave bands by the LP is computed, and then a DFB is applied to each band-pass Channel [69]

### 3.7.2 Why Contourlets?

There are several well-known systems that provide multi-scale and directional image representations. These include; 2-D Gabor wavelets [74], the steerable pyramid [75], 2-D directional wavelets [76] and brushlets [77]. However, the difference between these techniques and Contourlet transforms is that they do not allow for a different number of directions at each scale while achieving nearly critical sampling. In addition, the proposed construction employs iterated filter banks, which makes it computationally efficient [69].

One of the distinguishing features of contourlet transform is that they are defined on rectangular grids and thus offer a seamless translation to the discrete world, where image pixels are sampled on a rectangular grid. To achieve this ‘digital-friendly’ feature, the contourlet kernel functions [69] have to be different for different and cannot be obtained by simply rotating a single function. This is a key difference between the contourlet and the curvelet systems in [61, 65].
3.8 SUPPORT VECTOR MACHINES (SVM)

Support Vector Machines (SVMs) first introduced in 1992 by B.E.Boser et.al [78], are a set of related supervised learning methods used for classification. It is a relatively new classifier and is based on strong foundations from the broad area of statistical learning theory [79]. It belongs to a family of generalized linear classifiers. A special property of SVM is that it simultaneously minimizes the empirical classification error and maximizes the geometric margin; hence also known as maximum margin classifier [80]. It has found applications in a wide range of pattern recognition problems: handwritten character recognition, image classification, face detection, vehicle type recognition, bioinformatics, biomedical, signal analysis, medical diagnostics and in data mining.

SVM has become, in practice, the classifier of choice of numerous researchers and practitioners for several real-world classification problems. This is because SVM is capable of generalizing well (predicting the unseen or unknown samples with a good degree of accuracy) as compared to many traditional classifiers (Neural Network).

3.8.1 Basic Theory of SVM

Linear Decision boundary: SVM was originally developed for two class classification problem and later on extended to the multi-class case. SVMs perform pattern recognition between two classes by finding a linear decision surface that can separate the two classes.
For data which are linearly separable, there exists a large number of separating hyperplanes (see Figure 3.19(a)-(d)). SVM tries to construct a separating hyperplane, which maximizes the distance between the hyperplane and the nearest data point of each class (see Figure 3.19(e)).

**Optimal Hyperplane:** The input to a SVM algorithm is a set \((x_1, l_1), \ldots, (x_n, l_n)\) where \(x_i \in \mathbb{R}^n, l_i \in \{-1, +1\}\) of labelled training data, where \((x_1, l_1), \ldots, (x_n, l_n)\) is the data and \(l_i \in \{-1, +1\}\) is the label. The training data is said to be linearly separable if there exists a vector \(J\) and scalar \(b\) such that the following inequalities are valid:

\[
J.x_i + b \geq 1 \text{ if } l_i = 1 \\
J.x_i + b \leq -1 \text{ if } l_i = -1
\]  

\((3.43)\) \hspace{1cm} \((3.44)\)
These inequalities can be written as

\[ I_i(Jx_i + b) \geq 1, \quad i = 1, 2, ..., n \]  

(3.45)

A linear decision surface/hyperplane equation is

\[ Jx + b = 0 \]  

(3.46)

The optimal hyperplane is the one which determines the direction \( \frac{J}{|J|} \) along which the distance between the projections of the training vectors of two different classes is maximal [82] (see Figure 3.20).

![Diagram](image)

**Figure 3.20** An example of separable problem in a 2 dimensional space. The support vectors, marked with gray squares define the margin of largest separation between two classes [37].

*Computing the distance/margin width (M):* Suppose \( x^- \) be any point on the minus plane and \( x^+ \) be the closest point to the \( x^- \) on the plus plane. So to get from \( x^- \) to \( x^+ \), some distance is travelled in the direction of \( J \) (line from \( x^- \) to \( x^+ \) be perpendicular to the planes):

\[ x^+ = x^- + \gamma J \]  

(3.47)
\[ x^+ - x^- = \gamma J \]
\[ M = |x^+ - x^-| = |\gamma J| \]  
(3.48)

From (3.43) and (3.44), it is also known that
\[ J.x^+ + b = +1 \]  
(3.49)
\[ J.x^- + b = -1 \]  
(3.50)

Substituting value of \( x^+ \) from (3.47) in to (3.49)
\[ J.(x^- + \gamma J) + b = 1 \]
\[ J.x^- + b + \gamma JJ = 1 \]  
(3.51)
\[ -1 + \gamma JJ = 1 \]
\[ \gamma = \frac{2}{JJ} \]  
(3.52)

Equation (3.48) becomes
\[ M = \gamma|J| = \frac{2\sqrt{JJ}}{JJ} = \frac{2}{\sqrt{JJ}} \]  
(3.53)

Hence the hyperplane that optimally separates the data is the one that minimises
\[ \varphi(J) = \frac{1}{2}\|J\|^2 \]  
(3.54)
under the constraint (3.45). Constructing an optimal hyperplane is therefore a Quadratic Programming (QP) problem. The solution to the optimization problem of (3.54) under constraint (3.45) is explained in [82] using QP.

According to [82], the vector $J$ can be written as the linear combination of support vectors:

$$J = \sum_{i=1}^{N} \alpha_i I_is_i$$

Where $s_i$ is a set $N$ of support vectors, $\alpha_i$ is coefficient weights, $I_i$ is class labels of the support vectors. Further information can be found in [82, 83].

**Soft Margin Hyperplane:** So far the discussion has been restricted to the case where the training data are linearly separable. Consider the case where training data are not separable without errors. In such a situation, one may want to separate data using minimal errors. To generalise the optimal hyperplane to a non-separable case, slack variables $\varepsilon_i$ are introduced [82]. Constraint of (3.45) is now modified as:

$$I_i(J.x_i + b) \geq 1 - \varepsilon_i \quad \varepsilon_i \geq 0, \quad i = 1, 2, ..., n$$

(3.56)

The generalised optimal hyperplane is determined by minimizing

$$\varphi(J, \varepsilon) = \frac{1}{2} \|J\|^2 + \delta \sum_{i=1}^{N} \varepsilon_i$$

(3.57)

Where $\delta$ is a given value subject to constraint (3.56).

The solution to the optimization problem is found by following a procedure similar to that of determining optimal hyperplane using QP [82].
Extension to non-linear decision boundary: The optimal hyperplane discussed in the previous section has a linear decision boundary. This can be generalised to the non-linear case where a mapping function $\phi(x_i)$ is used to map the $n$-dimensional input vector $x$ into $N$-dimensional (higher dimensional) feature vector such that the non-linear hyperplane becomes linear (See Figure 3.21):

![Feature mapping diagram](image)

**Figure 3.21:** Example of Feature mapping. (Left) input space (right) feature space [82].

$$\phi : \mathbb{R}^n \rightarrow \mathbb{R}^N$$

$$\phi(x_i) = \phi_1(x_i), \phi_2(x_i), ..., \phi_N(x_i), \quad i = 1, ..., n \quad (3.58)$$

An $N$ dimensional linear separator $w$ and scalar $b$ is then constructed for the transformed vectors (3.58).

Classification of an unknown vector $x$ is done by first transforming it into feature space ($x \mapsto \phi(x)$) and then taking the sign of the function

$$f(x) = J \cdot \phi(x) + b \quad (3.59)$$

According to [82] as given in equation (3.55), one can write $J$ as a linear combination of support vectors in the feature space as:

$$J = \sum_{i=1}^{n} \alpha_i \phi(x_i) \quad (3.60)$$
Substituting value of $J$ from (3.60) into (3.59)

$$f(x) = \phi(x)J + b = \sum_{i=1}^{n} \alpha_i \phi(x_i) \phi(x) + b \quad (3.61)$$

The general forms of dot-product in Hilbert space [84] is:

$$\phi(u) \phi(v) = K(u,v). \quad (3.62)$$

Convolution of the dot-product in feature space can be given by a function satisfying Mercer’s Theorem [85]. These functions are called Kernel functions. Using different dot products $K(u,v)$, one can construct different learning machines with arbitrarily type of decision surfaces [86]. Some of the most popular Kernel functions are:

- Polynomial kernel with degree $d$
  $$K(u,v) = (u^T v + 1)^d$$

- Radial basis function kernel with width $\sigma$
  $$K(u,v) = \exp(-\|u - v\|^2 / (2\sigma^2))$$

The decision surfaces of these learning machines has the form

$$f(x) = \sum_{i=1}^{n} y_i \alpha_i K(x,x_i) \quad (3.63)$$

Where $x_i$ is a support vector in the input space and $\alpha_i$ is the weight of support vector in the feature space. To find the vectors $x_i$ and weights $\alpha_i$ one follows the same solution as for optimal hyperplane and soft margin hyperplane with some difference in calculation of matrix D [82].
3.8.2 Advantages of SVM

SVM offers several advantages which are typically not found in other classifiers [83, 87]:

- Computationally much less intensive (especially in comparison to Neural Networks).
- SVM is a universal machine. By changing kernel function $k(u, v)$ for convolution of dot-product, one can implement different networks [82].
- Lack of training data is often not a severe problem.
- Performs well in higher dimensional spaces (a factor which limits many efficient classifiers).
- Robust with noisy data (noise can severely degrade the performance of a Neural Network).
- Does not suffer as much from the curse of dimensionality and prevents overfitting (i.e. when a learning algorithm performs quite well on the training set, compared to its true performance on unseen test data).

3.9 CROSS VALIDATION

The theory of cross validation was invented by Seymour Geisser [88]. According to [89-90], cross-validation[91-93] is one of several approaches to estimating how well the model which has been learned from some training data is going to perform in future for unseen data. It is the statistical practice of partitioning a sample of data into subsets such that the analysis is initially performed on a single subset, while the other subset(s) are retained for subsequent use in confirming and validating the initial analysis. The initial subset of data is called the training set; the other subset(s) are called validation or testing sets.
There are several types of cross-validations. One of the most important is discussed in the following section, i.e. section 3.9.1. Details on other types can be found in [93].

### 3.9.1 n-fold Cross Validation

In n-fold cross validation, the original sample is partitioned into \( n \) subsamples. Of the \( n \) subsamples, a single subsample is retained as the validation data for testing the model, and the remaining \( n - 1 \) subsamples are used as training data. The cross-validation process is then repeated \( n \) times (the folds), with each of the \( n \) subsamples used exactly once as the validation data. The \( n \) results from the folds then can be averaged (or otherwise combined) to produce a single estimation.

The advantage of this method is that all observations are used for both training and validation, and each observation is used for validating exactly once [90].

### 3.10 k-NEAREST NEIGHBOUR CLASSIFIER (k-NN)

In pattern recognition, the k-NN [94] is a supervised learning algorithm for classifying objects based on a category of the majority of closest/neighbour \( k \) training examples \((k>0)\) in the feature space [95]. It is simplest of all machine learning algorithms. A number of different types of nearest neighbour finding algorithms exists [95]. These include:

- Linear Scan
- Kd-trees
- Locality sensitive hashing (LSH)
3.10.1 \textit{k-NN} Algorithm

The training examples are represented by position vectors in a multidimensional feature space. The space is partitioned into regions by locations and labels of the training samples. A point in the space is assigned to the class $c$ if it is the most frequent class label among the $k$ nearest training samples. Euclidean distance \cite{96} is usually used for this purpose.

During the training phase of the algorithm, the feature vectors along with their class labels are stored for all training samples. When a new test sample comes in (whose class is unknown), it is first represented by a feature vector in feature space similar to those of the training samples. To classify the test sample, the distance of the test feature vector to all stored training feature vectors is computed and $k$ closest samples are selected. An Example of \textit{k-NN} is illustrated in Figure 3.22.

\begin{figure}
\centering
\includegraphics[width=0.5\textwidth]{k-NN_example.png}
\caption{The test sample (green circle) should be classified either to the first class of blue squares or to the second class of red triangles. If $k = 3$ it is classified to the second class because there are 2 triangles and only 1 square inside the inner circle \cite{95}.}
\end{figure}

The test vector can be classified to a particular class by number of ways; one of the most used techniques is to classify the test vector to a class which is the most
common amongst the K nearest neighbours. A major drawback to using this
technique is that the classes with the more frequent examples tend to dominate the
prediction of the test vector, as they tend to come up in the K nearest neighbours
when the neighbours are computed due to their large number. One of the ways to
overcome this problem is to take into account the distance of each K nearest
neighbours with the new vector that is to be classified and predict the class of the
new vector based on these distances [95].

3.10.2 Strengths and Weaknesses of k-NN Classification

Strengths:
- Robust to noisy training data
- Effective if training data is large

Weaknesses:
- Need to determine the value of parameter K (number of nearest neighbours).
- Computation cost is quite high because one needs to compute distance of
each query instance to all training samples. Some indexing (e.g. K-D tree)
may reduce this computational cost

3.11 BOOSTING

‘Boosting refers to a general and effective method of producing a very accurate
prediction rule by combining rough and moderately inaccurate rules- of- thumb’
[97]. It gradually improves the accuracy of a learning algorithm by concentrating on
the ‘hardest’ examples (those most often misclassified by previous rules of thumb)
over each round and combines these weak prediction rules in to a single strong
prediction rule by taking the (weighted) majority vote of these weak prediction
rules.
CHAPTER 3

3.11.1 Adaptive Boosting (AdaBoost)

The Adaptive Boosting (AdaBoost) algorithm is a boosting algorithm first introduced by Freund and Schapire [98] in 1995. It combines many low-accuracy classifiers (weak learners) to create a high-accuracy classifier (strong learners). It is an adaptive algorithm to boost a sequence of classifiers, such that the weights are updated dynamically according to errors in previous learning. AdaBoost has been used as a classifier and feature selection technique in many applications [45, 46] [99-104].

3.11.2 AdaBoost Algorithm

According to Y. Freund and R. Schapire [97, 98], pseudocode for boosting is:

- Given: a training set of \((x_1, l_1), \ldots, (x_m, l_m)\) where \(x_i \in X\) are the instances of some domain \(X\), and \(l_i \in L = \{+1, -1\}\) are labels of the instances. (Note that two-class case is discussed here).
- A distribution of weights \(D_{i, \epsilon}\) is maintained over the training set of \(m\) samples. Initially all training samples are given equal weights i.e.
  \[ w_i = D_{i, \epsilon}(i) = \frac{1}{m} \text{ for } i = 1, \ldots, m. \]
- Call a given weak learning algorithm repeatedly in a series of \(R\) rounds. On each round, the weights of incorrectly classified examples are increased whereas those of correctly classified are decreased. The purpose is to make weak learner to focus on the hard examples (incorrectly classified) in training set. The weak learner's job is to find a weak hypothesis/classifier \(\hat{h}\), with a small error \(\epsilon\), on the distribution \(D_{i, \epsilon}\). The error is measured with respect to the distribution \(D_{i, \epsilon}\), on which the weak learner is trained. Once weak
CHAPTER 3

A hypothesis is obtained, a parameter $\alpha$, is calculated which measures the importance assigned to $\hat{h}_r$. Mathematically it is written as:

For $r = 1, \ldots, R$ :

- Set $Dis_r = \frac{w'}{\sum_{i=1}^n w'_i}$
- Train weak learner on distribution $Dis_r$.
- Find a weak hypothesis/classifier $\hat{h}_r : X \rightarrow \{+1,-1\}$ with small error

$\varepsilon_r = \sum_{i=1}^m Dis_r(i) | \hat{h}_r(x_i) - I_i |$

- Select $\alpha_r = \frac{1}{2} \ln \left( \frac{1 - \varepsilon_r}{\varepsilon_r} \right)$
- Update

$$Dis_{r+1}(i) = \frac{Dis_r(i)}{Z_r} \times \begin{cases} e^{-\alpha_r} & \text{if } \hat{h}_r(x_i) = I_i \\ e^{\alpha_r} & \text{if } \hat{h}_r(x_i) \neq I_i \end{cases}$$

$$= \frac{Dis_r(i) \exp(-\alpha_r, I_i \hat{h}_r(x_i))}{Z_r}$$

Whereas, $Z_r$ is a normalization factor

- Output the final hypothesis/classifier which is a weighted majority vote of the $R$ weak hypothesis.

$$\hat{H}(x) = \text{sign} \left( \sum_{r=1}^R \alpha_r \hat{h}_r(x) \right)$$

AdaBoost is typically used to solve two-class problems. It can be extended to multi-class classification problem [98,105]. Further details can be found in [98].
3.11.3 Advantages of AdaBoost

The AdaBoost has the following advantages:

- It is fast, simple and easy to program.
- It has no parameters to tune (except for the number of round).
- It requires no prior knowledge about the weak learner and so can be flexibly combined with any method for finding weak hypotheses.
- A nice property of AdaBoost is its ability to identify samples that are either mislabelled in the training data, or which are inherently ambiguous and hard to categorize. Because AdaBoost focuses its weight on the hardest examples, the examples with the highest weight often turn out to be outliers [97].

3.12 DATASETS

One important prerequisite of experimentation is the collection of a car image dataset and preparation of an associated database. From the literature survey it is observed that most of the literature only deals with images consisting of either the frontal or the rear views of cars. Apart from maintaining simplicity an important reason for this selection (rather than using sides or 3D views) is that front and/or rear profiles provide richer and thus most discriminating features. Therefore all test images used in experiments have been limited to car frontal views.

Two datasets were originally prepared/obtained for the use in the experiments.

**Dataset-1:** The images of this dataset have been collected from various car parks inside Loughborough University UK over a period of one month in the year 2006. The database consists of 25 different classes (Note that each class represents a unique make-model combination giving particular attention to the change of appearance of a certain model depending on the year of manufacture, e.g. a BMW-3 series registered in 1995 will belong to a different class when compared to one
registered in 2001). A total of 300 images of resolution 2048x1536 pixels were collected. The images were captured using a hand-held (note: this may cause variations in both the scale and in-plane rotation of the vehicles in the images) digital camera approximately 3 meters in front of the cars. For ease of reference the classes have been listed in Table 3.1.

<table>
<thead>
<tr>
<th>Vehicle Classes</th>
<th>Audi a4</th>
<th>Citroen xsara</th>
<th>Ford ka</th>
<th>Honda civic</th>
<th>Vauxhall astra</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bmw-series3</td>
<td>Ford mondeo</td>
<td>Peugeot 406</td>
<td>Volkswagen passat</td>
<td>Vauxhall corsa</td>
<td></td>
</tr>
<tr>
<td>Bmw-series5</td>
<td>Ford fiesta</td>
<td>Rover 45</td>
<td>Rover 25</td>
<td>Vauxhall vectra</td>
<td></td>
</tr>
<tr>
<td>Citroen ax</td>
<td>Ford ka</td>
<td>Peugeot 306</td>
<td>Toyota yaris</td>
<td>Volkswagen golf</td>
<td></td>
</tr>
<tr>
<td>Peugeot 206</td>
<td>Ford focus</td>
<td>Rover 75</td>
<td>Toyota corolla</td>
<td>Volkswagen polo</td>
<td></td>
</tr>
</tbody>
</table>

The classification of the car images requires the use of some images for training and the rest for testing. In experiments, generally, for each class, 8 training images and 2-4 test images have been used. Cross validation of samples was used to maintain fairness of experiments. Further to avoid considering the background areas of the images which are often easily captured while photographing a car, image cropping has been used, to separate out a region that just contains all features of the car frontal view from the background clutter. This was achieved following the cropping procedure [23] that was discussed in section 2.2.1 of chapter 2, which is based on first detecting and locating the license plate.
Dataset-2: The second dataset was obtained from a large database of car frontal images made available by V. Petrović and T. Cootes [23]. Each image was of resolution 640x480 pixels. The images were cropped following the algorithm [23] explained in section 2.2.1 of chapter 2, based on licence plate locations separately provided in the database alongside individual images. Table 3.2 illustrates the selected classes.

Figures 3.23 and 3.24 illustrate sample images from both datasets.

<table>
<thead>
<tr>
<th>Classes</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Audi a4</td>
<td>Fiat punto</td>
<td>Ford ka</td>
<td></td>
<td></td>
<td>Vauxhall astra</td>
</tr>
<tr>
<td>Bmw3</td>
<td>Fiat punto new</td>
<td>Ford mondeo new</td>
<td></td>
<td></td>
<td>Vauxhall astra new</td>
</tr>
<tr>
<td>Bmw5 new</td>
<td>Ford fiesta</td>
<td>Honda civic new</td>
<td>Rover 25</td>
<td></td>
<td>Vauxhall vectra</td>
</tr>
<tr>
<td>Citroenax</td>
<td>Ford fiesta new</td>
<td>Peugeot 306 new</td>
<td>Toyota Yaris</td>
<td></td>
<td>Volkswagen golf 3</td>
</tr>
<tr>
<td>Fiat Brava</td>
<td>Ford Focus</td>
<td>Renault 19</td>
<td>Toyota Corolla</td>
<td></td>
<td>VolksWagen Polo</td>
</tr>
</tbody>
</table>

Table 3.2: subset of 25 classes from car database of [23].
Figure 3.23: Car images taken for preparation of database from Loughborough University.
It was mentioned that cropping of images will require an initial detection and the subsequent identification of the location of the license plate, as the cropped area is decided based on the location and size of the licence plate. The images of dataset-2 originally consisted of licence plate information and therefore were more frequently used in our experiments. Dataset-1 has been provided in the public domain [115] for the wider benefit of the research community.

3.13 SUMMARY

This chapter has presented the fundamental theory upon which the contributions of this thesis (see Chapters 4-7) have been built. The chapter aimed at providing sufficient theoretical background to the readers so as to enable the understanding of
the theories used in the contributory chapters. Readers interested in further details have been referred to the original, related publications. The chapter also provided details of the test image databases that have been used in the research presented in the thesis.
CHAPTER 4

Use of Two Dimensional Statistical Linear Discriminant Analysis (2DLDA) in Vehicle MMR

4.0 INTRODUCTION

This chapter presents a Two Dimensional Linear Discriminant Analysis (2DLDA) [54, 55] based appearance based vehicle MMR approach. 2DLDA has been previously successfully used in face recognition [37, 38, 39, and 55]. The novelty of this work is its application to vehicle MMR where the general differences between the datasets of application domains face recognition and vehicle MMR provided in section 2.3 (see chapter2) requires specific design considerations to be followed. Further in the past literature a detailed comparison between the effects of selectively using eigenvectors when using 2DLDA has not been presented. This has lead to a lack of understanding about how to use Linear Discriminant Analysis effectively under varying illumination conditions and partial occlusion. The research presented in this chapter contributes to the state-of-art by addressing this issue.

For clarity of presentation, this chapter is divided in to several sections. Section 4.1 discusses the motivation behind this work; Section 4.2 discusses the proposed algorithm, section 4.3 provides experimental results and a comprehensive analysis of the performance of the proposed algorithm, whilst section 4.4 summarises and concludes the chapter.
4.1 MOTIVATION

A Principal Component Analysis (PCA) [48] (see chapter 3) based vehicle MMR approach has been previously proposed in literature [23]. In this work, it was observed that the negative effects on recognition accuracy due to the functional limitations of feature detectors under adverse lighting and occlusion conditions can be removed by applying PCA on the detected features [49, 52]. Within the present proposed research context, a modified version of the above PCA based approach has been used as a benchmark algorithm. Instead of applying PCA on a specific feature set (e.g., Edge Orientation, Direct Normalized Gradients, Harris Corner Response, Square mapped Gradients) of the images, PCA is applied directly on the image pixel data. This allows a fair comparison between the use of PCA and the proposed 2DLDA based approach. Preliminary investigation revealed that when used in a vehicle MMR context, the general increase of the overall scatter between images solicited by PCA can result in misclassification. It is observed that in the identification of vehicles (i.e. cars within this research context) the original data set can be separated into several groups, depending on make/model. Using PCA for recognition in such a scenario increases the within class scatter. Therefore two cars with the same make/model properties, but with slight illumination changes, can be scattered further apart and classified as being of different make/model. Thus PCA is good from data representation point of view but not suitable for data discrimination (See section 4.3 and [52] as a proof).

On the other hand, Linear Discriminant Analysis (LDA) by R.A Fisher [50] maximizes the ratio of between-class variance to the within-class variance in any particular data set thereby maximizing separability. Unfortunately for a high dimensional and small sample size problem such as vehicle MMR, the traditional LDA encounters two aspects of difficulties. Firstly, it cannot be used in cases where within-class scatter matrix is always singular. Secondly, the high dimensional image vectors lead to computational difficulty: LDA as proposed in [50] is based on
analysis of vectors. Based on these vectors the covariance matrix is calculated and an optimal projection is obtained. However, typical images are high dimensional patterns and therefore results in a high-dimensional vector space, where the evaluation of the covariance matrix is computationally costly.

2DLDA first proposed by Ming Li [55] (see chapter 3) provides effective solutions to the above problems of traditional one-dimensional LDA by directly extracting features from an image matrix, rather than a feature vector created out of an image to compute the between-class scatter matrix and the within class scatter matrix. This way evaluation of covariance matrix becomes easy. In the proposed work 2DLDA has been adopted for robust and efficient vehicle MMR. Further details of 2DLDA could be found in [55]. It is noted that neither the use of LDA nor the use of 2DLDA in vehicle MMR has previously been investigated in the literature.

4.2 Proposed Method

Figure 4.1 illustrates the block diagram of the proposed 2DLDA based approach to vehicle MMR.

By replacing the 2DLDA block by PCA, PCA based vehicle MMR approach is obtained that has been used as a benchmark to evaluate the 2DLDA approach and its performance efficiency. It should be noted that the database of images of cars used in the experiments represents only the frontal view of cars. Further these images have been captured under different lighting and weather conditions and from an approximately similar distance and height.
CHAPTER 4

Training Images

Test Image I

Extracted car font

Normalized test
Image I

Normalized training Images

Database

Feature matrices of training images
i.e.

[Y1, Y2, ..., Yn]

Image size normalization

2DLDA

Projection matrix

Wopt

Feature matrix

Y for
test image I

Y = IWopt

Distance Calculation

D(Y, Y1)
D(Y, Y2)
...
D(Y, Y3)

Minimum Distance

Min (D)

Figure 4.1: Proposed, 2DLDA based Vehicle MMR System.
The operation of each block of Figure 4.1 can be detailed as follows:

4.2.1 Vehicle Front Extraction and Normalization

Images in the database may contain background areas which potentially complicates their computer based processing. Therefore corresponding Region-Of-Interest (ROI) need to be extracted, that contain visually significant features which can be used in effectively distinguishing between different makes and models of cars. These ROIs are the frontal views of the car. After extraction of the ROIs, the images are normalized to the same resolution. These stages can be summarized as follows:

**Locating the ROI:** Given the number plate coordinates [23] (top-left and bottom-right corners), the number plate detection and relative ROI identification / measurement strategy proposed by V.Petrović and T.Cootes [23] and explained in section 2.2.1 (see chapter 2) has been adopted in the proposed approach.

**Image Normalization:** Cropping the ROI as explained above results in images of different size/scale. Therefore initially a size normalization of these images is carried out with the use of the 'imresize' function of MATLAB. This is subsequently followed by a 'bilinear interpolation' stage that removes any pixelization effects of the image resizing stage, resulting in smoother looking images. [Note: all images have been converted to grayscale before processing as 'colour' is insignificant in vehicle MMR].

4.2.2 Processing the Training Image Set

The cropped and normalized images of the training image set are first grouped manually according to their make and model. Subsequently following 2DLDA theory (see chapter 3), the optimal projection matrix $W_{opt}$ is obtained as follows:
• Assume that each training image is of dimensions $r \times c$. First of all, the Fisher projection axes are constructed by finding the orthonormal eigenvectors of $S_w^{-1}S_B$ corresponding to the first $m$ largest eigenvalues. [Note: $S_w$ and $S_B$ are calculated as detailed in chapter 3]. Then, eigenvectors corresponding to the largest eigenvalues (i.e. projection axes in the eigenspace) are used to obtain the optimal, Fisher projection matrix, $W_{opt} = [u_1, u_2, ..., u_m]$ with dimensions $c \times m$.

The matrix $W_{opt}$ is subsequently used for feature extraction. For a given image $I$, $Y_k = IW_k$ (where $k = 1, 2, ..., m$). These feature vectors are then placed in the form of a matrix $[Y = Y_1, ..., Y_m]$, i.e. the Fisher feature matrix of image $I$, with dimensions $r \times m$.

4.2.3 Processing the Test Images

Each test image initially undergoes cropping and normalization as described above for the images in the training set. Using $W_{opt}$ each normalized image is finally converted to a feature matrix in the eigenspace as described in section 4.2.2.

4.2.4 Classification

Classification of the test images into one of the given classes in the training set of car images is done by using $L_2$ norm metric. Euclidean distance [96] is calculated between the feature matrix of a test image $I$ and each of the projections in the data base of training images as follows:
• Given two images \( I_1, I_2 \) represented by the Fisher feature matrices, 
\[ y^1 = [y^1_1, \ldots, y^1_m] \text{ and } y^2 = [y^2_1, \ldots, y^2_m], \]
their overall Euclidean distance is defined as 
\[ d(y^1, y^2) = \sum_{i=1}^{m} \| y^1_i - y^2_i \|_2, \]
where \( \| y^1_i - y^2_i \|_2 \) denotes the Euclidean distance between the two Fisher feature vectors \( y^1_i \) and \( y^2_i \).

Finally, the make and model of the training image which gives the minimum distance to the test image is selected as the make and model of the test image.

### 4.3 EXPERIMENTAL RESULTS AND ANALYSIS

**Data for Experiments:** A car image database [23] of 200 images has been used which comprises of 25 different car make-model groups (see Table 3.2 chapter 3). All images are grayscale and have been cropped to a resolution of 70\( \times \)140 pixels using technique [23], to include an area around the head lights, upper and bottom grills.

**Experiment 1:** Car images with limited range of illumination variation, using all eigenvectors

To compare the performance of PCA and 2D LDA approaches under normal lighting conditions 71 test images were used. Within this experiment, the average illumination levels of the test images used were not largely different from those in the training set. View occlusions were also not present. All eigenvectors were considered in creating the Eigen and Fisher feature matrices.

**Results:** Overall analysis revealed that the 2D LDA based approach gave an identification accuracy of 87% as compared to the 78% accuracy obtained by the benchmark PCA based approach. The recognition accuracy was measured as a
percentage of the ratio of the number of times the best matching make and model being the correct match, to the total number of images tested for a given experiment.

A sample of experimental results is tabulated in Table 4.1.

Experiment 2: Car images with illumination variation and occlusion, using all eigenvectors

To analyze the relative performance of the 2DLDA and PCA approaches under varying illumination and occlusion conditions and to contrast with their performance under normal lighting and occlusion free conditions, a further experiment was carried out. A new set of 25 test images was constructed by altering a randomly selected sub set of test images used in the original test image set of 71. The alterations in the form of acute illumination changes and occlusion effects were introduced using Adobe Photoshop 7.0. The training image set used was not altered and was the same as in the previous experiment. Further, all eigenvectors were considered in creating the Eigen and Fisher feature matrices.

Results: A sample of the results is illustrated in Table 4.2. Overall analysis revealed that the 2DLDA based approach gave an identification accuracy of 76% as compared to the 48% accuracy obtained by the benchmark PCA based approach. It is observed that the 2DLDA based approach is able to correctly identify the car make and model even under large occlusion differences. Occlusions by both white and black objects were considered and the results were observed to be consistent for the 2DLDA based approach. It was further observed that when using the PCA based approach under both large variations in illumination and occlusion, the matching image found was mainly similar in average illumination level, rather than in terms of features. This is justifiable as the theoretical evaluation of the PCA based approach suggests that it follows an appearance based matching approach rather than a feature based matching approach.
Analysis: The analysis of the above experimental results leads to the conclusion that the 2DLDA based approach is more efficient and robust under varying illumination/lighting and occlusion conditions, when compared to the direct PCA based approach. Changing illumination and occlusion increases the variance amongst images. Unfortunately PCA further increases variance based on texture information, across all the images in the Eigen feature space, with disregard to any make-model groupings. This leads to potential misclassification. In contrast, in the Fisher feature space, despite increase in variance between images, 2DLDA minimizes within class variance while maximizing between class variance. However, despite the significantly better performance of the 2DLDA based approach as compared to the PCA based approach, it is seen that both approaches have suffered a noticeable loss of accuracy under the varying illumination and occlusion conditions.

Table 4.1: Comparison of 2DLDA vs. PCA under normal lighting conditions using all eigenvectors. Only cases where results defer have been illustrated.

<table>
<thead>
<tr>
<th>#</th>
<th>Test Image</th>
<th>2DLDA Result</th>
<th>PCA Result</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td><img src="image1.png" alt="Image 1" /></td>
<td><img src="image2.png" alt="Image 2" /></td>
<td><img src="image3.png" alt="Image 3" /></td>
</tr>
<tr>
<td>2</td>
<td><img src="image4.png" alt="Image 4" /></td>
<td><img src="image5.png" alt="Image 5" /></td>
<td><img src="image6.png" alt="Image 6" /></td>
</tr>
<tr>
<td>3</td>
<td><img src="image7.png" alt="Image 7" /></td>
<td><img src="image8.png" alt="Image 8" /></td>
<td><img src="image9.png" alt="Image 9" /></td>
</tr>
</tbody>
</table>
Table 4.2: 2DLDA vs. PCA under varying lighting and occlusion conditions considering all eigenvectors. Only cases where results defer have been illustrated.

<table>
<thead>
<tr>
<th>#</th>
<th>Test Image</th>
<th>2DLDA Result</th>
<th>PCA Result</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td><img src="image1.png" alt="Image" /></td>
<td><img src="image2.png" alt="Image" /></td>
<td><img src="image3.png" alt="Image" /></td>
</tr>
<tr>
<td>2</td>
<td><img src="image4.png" alt="Image" /></td>
<td><img src="image5.png" alt="Image" /></td>
<td><img src="image6.png" alt="Image" /></td>
</tr>
<tr>
<td>3</td>
<td><img src="image7.png" alt="Image" /></td>
<td><img src="image8.png" alt="Image" /></td>
<td><img src="image9.png" alt="Image" /></td>
</tr>
<tr>
<td>4</td>
<td><img src="image10.png" alt="Image" /></td>
<td><img src="image11.png" alt="Image" /></td>
<td><img src="image12.png" alt="Image" /></td>
</tr>
</tbody>
</table>

Experiment 3: car images without illumination variation and with illumination variation using reduced eigenvectors

To evaluate the performance of PCA and 2DLDA approaches when using a reduced feature space, further experiments were carried out. The results are tabulated in Table 4.3. Experiments were carried out for both approaches, when all eigenvectors, all but the eigenvectors with the three largest eigenvalues (hereafter called the three most significant eigenvectors) and all but the eigenvectors with the hundred smallest...
eigenvalues (hereafter called the hundred least significant eigenvectors), were considered. Further two sets of experiments were performed. One with the initial training set of 200 images where the luminance variances of make-model groups were maintained within a limited range and an updated training set obtained by replacing some images within each group of the initial training set, with images that have higher luminance variances from the rest of the images in the group. In creating the above updated training set, the number of images per group, was maintained. A total of 71 images were used for testing.

Results and analysis: The results in table 4.3 illustrate that the performance of the PCA based approach degrades when the updated training set with images of higher illumination variation is used.

Table 4.3: 2DLDA vs PCA results in a reduced vector space

<table>
<thead>
<tr>
<th></th>
<th>MMR Accuracy (%)</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Using All</td>
<td>Dropped the</td>
<td>Dropping the</td>
</tr>
<tr>
<td></td>
<td>Eigenvectors</td>
<td>three most</td>
<td>100 least</td>
</tr>
<tr>
<td>Initial training set</td>
<td>PCA</td>
<td>78</td>
<td>85</td>
</tr>
<tr>
<td></td>
<td>2DLDA</td>
<td>87</td>
<td>79</td>
</tr>
<tr>
<td>Updated training set</td>
<td>PCA</td>
<td>66</td>
<td>76</td>
</tr>
<tr>
<td></td>
<td>2DLDA</td>
<td>70</td>
<td>64</td>
</tr>
</tbody>
</table>

This is expected as an increase of data scatter in the original image domain will lead to a largely increased scatter in their eigenspace domain, increasing chances of misclassification. Further, the results in Table 4.3 illustrate that the accuracy of the PCA based approach increases when the three most significant eigenvectors are ignored. This is due to the fact that removing the most significant eigenvectors removes the consideration of illumination variances between cars of identical make-
model in matching, thereby reducing chances of misclassification. This reasoning is further supported by the fact that dropping the most significant eigenvectors has resulted in a better percentage improvement of accuracy (i.e., 66% to 76% as against 78% to 85%), when the updated training dataset was used. Note that the updated training dataset contains groups of images that have high variations in luminance. Therefore using a PCA based approach that considers all eigenvectors is bound to perform sub-optimally. It is also seen that removing the hundred least significant eigenvectors, only marginally degrades the accuracy of the PCA based approach. However further removal of low significance eigenvectors will reduce the accuracy as the discrimination ability of the PCA is based on these eigenvectors.

The results in Table 4.3 also reveal that when using the 2DLDA approach dropping the hundred least significant eigenvectors has resulted in a considerable improvement of recognition accuracy in contrast to the behavior of the PCA based approach. It is further noted that the percentage improvement of accuracy obtained in the experiment where the updated training set is used (70% to 87%) is significantly more than where the initial training set is used (87% to 91%). These observations can be supported by the following theoretical reasoning: The least significant eigenvectors when using the 2DLDA approach signifies instances of low, between class scatter to within class scatter ratio. The presence of cars that are identical in make-model, but differ significantly otherwise due to the presence of illumination variations or occlusions, directly results in these eigenvectors. Ignoring these in matching therefore means ignoring the effects due to illumination variations and occlusions, which in turn positively impacts the recognition accuracy. The extra improvement observed above, when using the updated training set supports this argument. When the three most significant eigenvectors are dropped the 2DLDA based approach behaves in contrast to the PCA based approach. The recognition accuracy is decreased. This can be supported by the fact that in the fisher feature space, the highly significant eigenvectors refer to instances where the ratio, between class scatter to within class scatter is high. Ignoring these eigenvectors can therefore directly lead to misclassification.
In summary it can be stated that in general the 2DLDA based algorithm exceeds the performance of the PCA based approach. In particular the 2DLDA outperforms the PCA based approach under varying illumination and occlusion conditions. The best performance with the PCA based approach is obtained when the more significant eigenvectors are dropped, whereas the best performance with the 2DLDA is when the least significant eigenvectors are dropped.

It is further noted that once the training is completed, i.e. for example in the 2DLDA approach, when the Fisher feature matrix is calculated using the database of training images, testing for the make-model of a new car can be performed real time.

4.4 SUMMARY AND CONCLUSION

In this chapter a novel 2DLDA based approach to vehicle MMR has been proposed. Performance of the proposed method with a direct PCA based approach has been compared here. Experiments were designed and carried out to compare and contrast the two approaches under normal illumination conditions, adverse illumination variations and in the presence of occlusions. The results concluded that in general the 2DLDA performs better than the PCA in vehicle MMR. In particular the 2DLDA approach outperforms the PCA approach under varying illumination and occlusion conditions. Further detailed experiments have been provided to analyze the performance of the two algorithms when only a sub-set of eigenvectors are considered. It has been shown that the best performance with the 2DLDA approach is obtained when the eigenvectors of lower significance are ignored in contrast to the improvement obtained in the PCA approach when the eigenvectors with the highest significance are ignored. For the given database of 200 car images of 25 different make-model classifications, a best accuracy of 91% was obtained with the 2DLDA approach. The best accuracy obtained by the PCA approach was 85%.
CHAPTER 4

Though appearance based approaches like 2DLDA have been successful in recognition of vehicle types, it is limited in its ability due to scale, translation and rotation invariance. In practical situations, one may come across such problems. In an attempt to address these problems and the limitations of 2DLDA, in chapter 5 a further novel technique for vehicle feature extraction based on Scale Invariant Feature Transform (SIFT) has been proposed. Further, a novel matching strategy has been proposed.
CHAPTER 5

Use of Scale Invariant Feature Transforms (SIFT) in Vehicle MMR

5.0 INTRODUCTION

This chapter presents a novel approach for the use of Scale Invariant Feature Transforms (SIFT) [58, 59] (see chapter 3) in vehicle MMR. SIFT provides localised interest points in images, i.e. feature/keypoints, which are invariant to scale, rotation and illumination. It is noted that SIFT has been previously used as a feature extraction technique in vehicle MMR by L. Dlagnekov [25] and face recognition [40, 41] (see Chapter 2). However, a review of this work revealed that after the SIFT features are detected, the feature matching strategy adopted is ineffective and is not suitable for vehicle MMR applications (see section 5.1). Therefore in this chapter a novel approach to SIFT feature extraction and matching has been proposed that significantly improves recognition accuracy and robustness (see section 5.2) of its use in vehicle MMR.

For clarity of presentation, the chapter is divided into several sections. Section 5.1 discusses the motivation behind this work; Section 5.2 discusses the proposed algorithm, section 5.3 provides experimental results and a comprehensive analysis of the performance of the proposed algorithm, whilst section 5.4 summarises and concludes the chapter.

5.1 MOTIVATION

A straightforward approach to vehicle MMR using SIFT is to detect keypoints for all images in the database of cars and of the test images and to subsequently match the descriptors of the keypoints of the test image with those of the database images.
under specified nearest neighborhood criteria [25]. Under this approach a local image descriptor at each keypoint of the test image is matched against similar descriptors of all keypoints of each training image. If the matching error is below a specified threshold the keypoints are considered to be matching. Finally the database image, with which most number of test image keypoints are matched, is considered to represent the test image’s make and model. However a critical review of this approach revealed that two cars of the same make and model may not always reveal keypoints at corresponding locations (see Figure 5.1 – top row). Therefore a significant number of keypoints will either be wasted (i.e. not considered in matching) or may match with a keypoint from the image used for matching, which is located at a non-corresponding position (see Figure 5.1 –bottom row). Although the later could be prevented by limiting the search for matching keypoints to be within the neighborhood of corresponding locations, i.e. window based search, this often results in the keypoint being classified as unreliable in matching as no matching keypoints are present within the specified area of the image to be matched. This eventually leads to the situation where the keypoint for which a match is to be found being totally ignored in the overall matching process. This means that a key feature of the image is ignored and hence could directly lead to mismatches. Further experiments revealed that the differences in keypoints identified in images representing cars of identical make and model is due to differences in camera capture and environmental conditions. Given the practical nature of the application domain, vehicle MMR algorithms should be developed to be robust to these variations.
Figure 5.1: The top row figure shows the extracted features for two car images of the same make and model (Ford Fiesta). Both cars have features which do not correspond. The bottom figure shows the matches between individual keypoints, when the straightforward SIFT based approach is used. Only two of the displayed features found a match, out of which only one is correct.

5.2 PROPOSED APPROACH

Figure 5.2 illustrates a block diagram of the proposed approach to SIFT based vehicle MMR. The various stages of the algorithm can be explained as follows:
CHAPTER 5

5.2.1 Pre-processing the Image Database

In this stage, SIFT [58] is used to locate all keypoints and the corresponding descriptors of all images in the database. The images are initially cropped to include only the front grill, lights and bumper area of all cars and are subsequently normalized following an approach identical to that of [23]. Subsequently, for each image of the database, the set of identified keypoint descriptors are stored. Each descriptor is of size 128 [58]. As the images in the database can be captured under controlled lighting and camera conditions or can be obtained via manufacturer’s showroom photos, the number of detected keypoints per database image will be...
optimized. Therefore for a given model it is expected that a significant proportion of distinct features will be captured in the form of descriptors. It is also noted that the number of keypoints, $K$, detected for each make-model will differ.

5.2.2 Processing the Test Images

The test image initially undergoes cropping and normalization, identical to that used in database image preparation. Given a database image to match, all of its detected keypoints are projected to the test image and corresponding points are noted as so-called matchpoints. Around each matchpoint a window of $n*\text{n}$ pixels is selected. For each of the $n*\text{n}$ pixel locations, a local image descriptor is obtained by following an approach identical to that followed in finding the SIFT keypoint descriptors of the database image. Thus each matchpoint and its associated $(n*\text{n}-1)$ surrounding pixels will form a descriptor array of size $K*(n*\text{n})$.

5.2.3 Matching

Each keypoint descriptor of each database image is now compared with the descriptors calculated at the matchpoint and the surrounding $(n*\text{n}-1)$ pixels of the test image. The descriptor that gives the minimum distance matching ($L_2$-norm) is considered the best match and the relevant distance is noted. Once the distances of all best matching, ‘keypoint descriptor – test image pixel descriptor’ distances are calculated, they are averaged to find a single overall matching distance between the database image that was matched and the test image.
CHAPTER 5

5.2.4 Classification

A simple approach for classification has been used, i.e., the make and model of the database image that gives the minimum average matching distance from the test image is considered to be the make and model of the test image.

The basic idea of the proposed approach is therefore to find the best possible matching point on the test image corresponding to each of the key points of each database image. Therefore it eliminates the possibility of important keypoints being totally ignored in matching. Note that in the proposed approach the \( n \times n \) pixels considered on the test image need not represent keypoints of the test image, but yet they will be matched based on local descriptor information. This is the main reason of the vastly improved accuracy of the proposed system (see section 5.3).

5.3 EXPERIMENTAL RESULTS & ANALYSIS

Data: Experiments were carried out on frontal images of a set of 28 cars representing 7 make-models. Each image was of size 160*100 pixels and of greyscale nature. For each model 2 images were used for testing.

Experimental Results: Table 5.1 tabulates the experimental results.
CHAPTER 5

Table 5.1: Experimental results.

<table>
<thead>
<tr>
<th>Test Images</th>
<th>Best match results within each training class</th>
<th>Classes for Training set</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Bmw3</td>
</tr>
<tr>
<td>Bmw3-1.jpg</td>
<td>190736</td>
<td>201375</td>
</tr>
<tr>
<td>Bmw3-1blur.jpg</td>
<td>6531*</td>
<td>182384</td>
</tr>
<tr>
<td>Bmw3-5.jpg</td>
<td>309899*</td>
<td>175720</td>
</tr>
<tr>
<td>Audi4-3-dot.jpg</td>
<td>193494</td>
<td>22417*</td>
</tr>
<tr>
<td>Audi4-5.jpg</td>
<td>156637</td>
<td>173310*</td>
</tr>
<tr>
<td>Ford-focus-5.jpg</td>
<td>152531*</td>
<td>197925</td>
</tr>
<tr>
<td>Ford-focus-6.jpg</td>
<td>202447</td>
<td>198779</td>
</tr>
<tr>
<td>Peugeot-3.jpg</td>
<td>174247</td>
<td>188379</td>
</tr>
<tr>
<td>Peugeot-6.jpg</td>
<td>177247</td>
<td>207923</td>
</tr>
<tr>
<td>Vw-golf-5.jpg</td>
<td>156458</td>
<td>191308</td>
</tr>
<tr>
<td>Vw-golf-6.jpg</td>
<td>137638</td>
<td>196139</td>
</tr>
<tr>
<td>Bmw5-5.jpg</td>
<td>80547</td>
<td>206536</td>
</tr>
<tr>
<td>Bmw5-6.jpg</td>
<td>99794*</td>
<td>208666</td>
</tr>
<tr>
<td>Ford fiesta-3.jpg</td>
<td>168679</td>
<td>189674</td>
</tr>
</tbody>
</table>

Values in the Table 5.1 marked with "*" represent the best match. Results in Table 5.1 show that a wrong match was found only in one out of the 14 test cases (i.e. Ford-focus-5.jpg with Bmw3). It is noted that the database consists of two examples of models of cars from the same make that looks somewhat similar. These are between BMW 3 and 5 models and Ford Fiesta and Focus models. Further the test image 2 (Bmw3-1blur.jpg) represents a test image that is the blurred version of the test image 1 (Bmw3-1.jpg). Despite the image been severely blurred it shows that it has been able to identify the accurate make and model. These results were obtained when the search for best matching points for keypoints was done within a window size of 5*5 pixels centred at the matchpoint.

Analysis: It is noted that the window base search adopted within the proposed design of the vehicle MMR system, allows accurate make and model identification under limited image displacements. Further the free descriptor matching coupled with windowed search allows matching under limited rotational variance (note: it was mentioned in chapter 3 that the use of SIFT allows scale invariance in...
matching). In addition the proposed approach only requires the descriptors of the keypoints of the database images to be stored and used in matching. This saves considerable memory space as full details of database images need not be stored for processing.

The experiments above were conducted on a subset of the full database of car frontal images that comprises of a total of 200 cars of 25 different makes-models. When the proposed idea was tested on the entire database of 200 cars, an accuracy figure of 90% was obtained.

5.4 CONCLUSION AND FURTHER WORK

In this chapter a novel, robust, approach to vehicle MMR, based on Scale Invariant Feature Transforms has been presented. The proposed approach identifies keypoints on all database images and attempts to find the best matching point for each keypoint of each database image, within a windowed search area of a given test image. The points in the test image need not be classified as keypoints to enable matching, which is the key aspect that results in an improved accuracy of matching of the proposed approach as against using the traditional SIFT based matching. Experimental results have been provided to prove the effectiveness of the proposed algorithm. The window base search allows accurate make and model recognition under limited image displacements. The free descriptor matching coupled with window search allows matching under limited rotational variance. Further, the use of SIFT allows scale invariance in matching. Further, the proposed approach only requires the descriptors of the keypoints of the database images to be stored and used in matching. This saves considerable memory space as full details of database images need not be stored for processing. Further extension to this work is proposed in next chapter of this thesis.
CHAPTER 6

Use of Adaptive Boosting in Feature Selection for Vehicle MMR

6.0 INTRODUCTION

This chapter proposes a novel approach to Scale Invariant Feature Transform (SIFT) [58] based vehicle MMR in which SIFT [58] features are initially investigated for their relevance in representing the uniqueness of the make and model of a given vehicle class using an Adaptive Boosting (AdaBoost) [97, 98] based feature selection technique. This is an extension to the SIFT based vehicle MMR approach proposed in chapter 5. Experimental results are provided to show that the proposed selection of SIFT features significantly reduces the computational cost associated with classification at negligible loss of the system accuracy. It is further proved that the use of more appropriate vehicle matching algorithms enables significant gains in the accuracy of the algorithm. Experimental results are provided to prove that a 92% accuracy rate can be achieved on a publically available database of car frontal views.

For clarity of presentation, this chapter is divided in to several sections. Section 6.1 provides a discussion on research motivation and highlights open research issues in vehicle MMR. Section 6.2 discusses the proposed methodology in detail. Section 6.3, provides results of a number of experiments performed to prove the effectiveness of the proposed approach. Finally section 6.4 concludes with an insight to future improvements.
CHAPTER 6

6.1 RESEARCH MOTIVATION

Vehicle MMR approaches proposed in the literature (see chapter 2) are based on an initial stage of feature detection, where these detected features are subsequently used in matching. The majority of these methods rely on edge maps as features. However, the pixel domain edge extractors used (e.g. Canny [108]) are limited in their performance and therefore fail in accurately capturing the smooth curves/contours which are an important part of most images. F.Kazemi et.al and S.Rahati et.al in [32, 33] provided the solution of this problem by proposing techniques for feature extraction in the transform domain. However, even the best edge extractor could fail to identify all edges that will be required in uniquely defining the make and model of a vehicle, in cases where the captured images of the vehicles are not clear, due to adverse lighting effects, occlusion and pose/scale variations etc. The SIFT based vehicle MMR approaches of [25, 34], promise to address some of the shortcomings of traditional feature based approaches. Specifically, SIFT based approaches enable the extraction of invariant features from images that result in more robust feature based matching under occlusion, scale and rotation invariance. Thus SIFT based approaches have been particularly used in object recognition, where the object being searched is immersed in background clutter. The basic SIFT based approach to vehicle MMR [25] was based on matching the keypoints of a query image to the keypoints of images in a database. One shortcoming of this simple approach is that keypoints from the background (i.e. outliers) of the query and database images may dominate the matching process thereby resulting in wrong matches. As a solution to this problem, [34] suggested the use of RANdom SAmping Consensus (RANSAC) [109] to separate outliers from inliers. However this approach involved the detection of the vehicle boundary area using edge/contour detectors and then using an iterative algorithm RANSAC. The accuracy of this is highly dependent on the accuracy of the segmentation of the object area and the iterative process makes the approach time consuming. Further the features used in MMR are not a true representation of the uniqueness of the make-model (see Figure 6.1) and rotation invariance is affected by this approach.
In order to resolve these problems a novel approach to SIFT based vehicle MMR is proposed in this chapter. The idea is based on the fact that humans are able to identify a given vehicle’s make-model based on a mental matching of each model’s unique features [117], such as the shape of the grill, badge, shape of lights etc. It is shown that after the keypoints have been found, AdaBoosting [97, 98], popularly used in feature selection, can be used to select features that are most representative of a given make-model enabling its use in vehicle MMR. Experimental results are provided to prove the effectiveness of the proposed algorithm.

6.2 PROPOSED APPROACH

Figure 6.2 illustrates the block diagram of the proposed approach to vehicle MMR. The basic idea is to match the keypoints of a query image against selections of unique and most representative feature sets selected from each make-model of cars.
The stages involved can be detailed as follows:

### 6.2.1 Dataset Preparation

Two sets of training images are collected. In the first set, images from the training and test sets are cropped to include only the front grill, lights and bumper area of all cars using the cropping approach proposed in [23]. The Second set consists of...
CHAPTER 6

frontal views of cars that include background clutter such as other cars, parking slot markings, tarred surfaces, lamp posts etc. The test/query images consist of cropped frontal views of images of cars without background.

![Figure 6.3: Examples of the database of images](image)

6.2.2 Feature Detection

As the first step of the proposed processing algorithm, the use of interest point/keypoint detecting technique: SIFT [58] has been investigated. SIFT defines interest points as minima and maxima of the difference of Gaussians that occur at multiple scales. SIFT is invariant to image translation, scaling, rotation, partial illumination changes and affine or 3D projections, allowing a consistent detection of features on images of cars.

In the proposed application of SIFT, keypoints from all training images of a make-model are pooled together. Similarly keypoints for the test images are detected. Figure 6.4 illustrates the detected SIFT features from two individual training images (Audi A4) and the projection of all pooled keypoints from all training images on a selected image of an Audi A4 car. It shows that the keypoints concentrate near the grill, lights, badge and front bumper areas.
6.2.3 Feature Selection

Images of cars in practical situations are assumed to be taken on streets or in parking lots (see Figure 6.3). This presents the problem of having a background scene in the image that can greatly affect the relevance of interest points that are detected. Within the present research context, the method proposed to eliminate outliers, i.e. the interest points not associated with a car, is to use AdaBoost [97, 98].

According to Y.Freund and R.Schapire [97], a useful property of AdaBoost is its ability to identify outliers, i.e., examples that are either mislabelled in the training data or which are inherently ambiguous and hard to categorize. These are thus called 'hard' points, whereas robust points are called 'easy' points. As AdaBoost
focuses its weight on the hardest examples, the examples with the highest weights often turn out to be outliers. Further according to B.Caprile [112], Entropy [113] is used as a measure for separating ‘easy points’ from ‘hard’ points based on weight values. Exploiting this, use of feature selection approach has been proposed which is based on weight values attached to the keypoints/interest points of the training images of a particular make-model.

During the training phase, keypoints from training images of a particular make-model are compared against the rest of the classes (i.e. a two class problem is considered where one make-model represents positive samples and rest, negative samples). This is repeated for all make-models (Note that a class has been introduced in training set, which only consists of typical background area, i.e. no cars). Weight values attached to all keypoints of a particular make-model up to certain number of iterations of boosting are recorded. On completion of every iteration, weights associated with keypoints are updated in order to focus the algorithm’s attention on the hard points. Entropy is then calculated for the stored weights as: The interval [0, 1] is partitioned into $\frac{\hat{p}}{p}$ subintervals of length $\frac{1}{\hat{p}}$, and the entropy value is computed as:

$$-\sum_{i=1}^{\hat{p}} f_i \log_2 f_i \quad (6.1)$$

Where $f_i$ is the relative frequency of weight values falling in $i$-th subinterval ($0 \log_2 0$ is set to 0). For experiments in proposed technique, $\hat{p}$ was set to 500. These entropy values are first sorted and keypoints with lowest $N$ percent of entropy values are subsequently selected as the valid features for a particular car make-model. These selected features are the best representative features of a car make-model as they have low uncertainty factor in being classified in the right class.
6.2.4 Interest Point Matching

Two methods for matching interest points have been investigated. The first is the original SIFT feature matching procedure proposed by Lowe [58]. In this approach the keypoints from the test/query image are matched against each of the selected keypoint from a particular model using the Euclidean distance [96] as a measure. The pair of interest points having the minimum Euclidean distance is considered to be matched. The model in the database with the highest inliers count will be labelled as being the best matching image to the query image, and the corresponding make-model category will be used to label the query image. The Second approach adopted for matching is based on the SIFT matching algorithm proposed as a contribution towards this thesis in chapter 5. In this approach SIFT descriptors [58] of the selected keypoints of every training image is compared with SIFT descriptors of points within a maximum likelihood area of the test images, centred at the point that corresponds location wise to the keypoint of the training image. It is noted that images are cropped and normalised before the above method is used for matching.

6.3 EXPERIMENTAL RESULTS & ANALYSIS

Two experiments were conducted to evaluate the performance of the proposed algorithm.

Data for Experiments: Two datasets have been used to conduct the experiments. The first dataset consisted of 50 images of cars (frontal views) belonging to 5 different classes. These images were cropped at the top (only) to remove the clutter in background due to foreign objects, particularly other cars. However, some background clutter is visible in the sides. This dataset was collected in order to prove that AdaBoost can be used to identify features unique to a given make-model (see Figure 6.5) in the presence of other key-points in the training images.
The second dataset consisted of 300 images of cars (frontal views) belonging to 25 different classes (see table 3.2 chapter 3 for details). Each training class consisted of at least 8 images of different cars belonging to the same make and model. [Note: the number of cars in each class was not equal]. The images were cropped [23] in all four sides (as appropriate) to remove background clutter.

**Experiment 1: Determining the number of iterations required for the AdaBoost Algorithm.**

Initially an experiment was designed and conducted on the second dataset to determine the number of iterations ‘$R$’ required for the AdaBoost algorithm [97, 98] (discussed in section 3.11.2 chapter 3) to obtain a stable result of classification. In this experiment descriptors of all keypoints of all BMW-3 cars were labelled to be of one class and all other descriptors from all other models were labelled to be of a second class, i.e. a binary classification problem was addressed. An accurate classification result was noted when either a true positive or a true negative result was obtained. Experiments revealed that after $R = 200$, there is only a negligible decrease in accuracy and the accuracy was held constantly at approximately 91% beyond this point. This is illustrated in Table 6.1. Note that in the experiments the Nearest Mean Classifier (NMC) [114] has been used as the weak learner due to its simplicity.] A similar level of accuracy was obtained for other make-models at similar number of iterations.

<table>
<thead>
<tr>
<th>No of iterations $R$</th>
<th>% classification error</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>35</td>
</tr>
<tr>
<td>50</td>
<td>15</td>
</tr>
<tr>
<td>100</td>
<td>10</td>
</tr>
<tr>
<td>150</td>
<td>9</td>
</tr>
<tr>
<td>200</td>
<td>9</td>
</tr>
</tbody>
</table>

Automatic Approaches Towards Vehicle Make and Model Recognition 110
Experiment 2: Determine the effect of applying the proposed technique for feature selection.

A second experiment was designed to determine the effect of applying the proposed technique for feature selection from a pool of features obtained from car images of a particular make-model. To better visualise the effect of applying the proposed technique in feature selection, idea is initially applied on the first dataset. As this dataset consists of areas from the background, it is useful in demonstrating the fact that AdaBoost will be able to separate features unique to each make-model from the feature points of the background areas. Further note that the keypoints of a class associated with low entropy measures indicate low uncertainty in classification and are thus best to be used in classification. Therefore by selecting the keypoints associated with the lowest entropy figures, only the most unique features of a car make-model will be identified. These can be subsequently used in vehicle MMR (see Figure 6.5).

It is obvious from the results illustrated in Figure 6.5 that most of the keypoints with lowest entropy values lie on the visually significant parts of the car and thus represent the inliers. Experiments have been performed to determine a suitable threshold value for the entropy and found that keeping 25% of lowest entropy gives the best accuracy.
Experiment 3: Application of proposed feature selection technique on second set of data.

A further experiment was conducted on the second set of data (database of 300 images) to obtain the overall classification accuracies. Keypoints whose entropy values were within the lowest 25% were selected for subsequent processing. Since the second dataset consists of cropped images (from all sides), the above feature selection process helps to separate keypoints of the class (i.e. make-model) which
are likely to be easily confused with keypoints belonging to other classes. In other words the feature selection strategy adopted will be able to identify unique feature points that are distinctive for each make-model.

The effectiveness of the proposed feature selection technique has been discussed with the help of three classes (i.e. make-models): Audi-A4, Toyota corolla, Citroen, as examples and illustrated in Figure 6.6. The distinctive areas of three classes are highlighted in Figure 6.6 (a), (e) and (i) with different colour rectangles. Projection of keypoints in absence of feature selection on a selected image from each class (see Figure 6.6 (b), (f) and (j)), displays the concentrated region. These concentrated regions in each class are a mixture of similar and unique keypoints to other classes. For example, keypoints in the area of the lights in each car are not unique as the descriptors could be very similar in these regions. Similarly, some keypoints could appear on the images of cars due to surface reflection or other deforms, thus useless in identification of a make-model (see keypoints on bonnet areas of Figure 6.6(f) and (j)).

In order to extract the most representative and unique keypoints for each class, proposed feature selection comes into play its role. A Comparison of feature selection using the highest 25% entropy values to that of lowest 25% for each of examples class is illustrated in Figure 6.6 (c),(g),(k) and (d),(h),(l) of Figure 6 respectively. It is obvious from these figures that the lowest 25% entropy based selected keypoints preserve the uniqueness of class in much better way than the highest 25% entropy values. Presence of keypoints in distinctive areas, highlighted in (a), (e) and (i) is higher in figures (d), (h) and (l) than figure (c), (g) and (k). In addition keypoints due to noise (reflections, deforms) have been pruned in this case.
Figure 6.6: Feature selection from top to bottom row: no features, all keypoints, 25\% highest entropy keypoints, 25\% lowest entropy keypoints, on example classes from left to right: Audi-A4, Toyota corolla and Citroen.

After the selection of keypoints that are able to best represent unique features of all make-models, the data is ready for testing.
Experiment 4: Matching of a selected set of features from training car images to that of test car image using the SIFT [58] keypoint matching approach.

This method uses the AdaBoost feature selection procedure proposed in this chapter, prior to the use of the original SIFT keypoint matching algorithm proposed in [58]. Matching results based on selected features with lowest 25% entropy are illustrated in Figure 6.7. Note the high degree of correspondence between the matching keypoints.

![Figure 6.7: Matching results.](image)

An accuracy of 82% classification was achieved when the proposed, AdaBoost based feature selection method was adopted; with the original SIFT keypoint matching scheme [58] as compared to 83% when all keypoints were considered in matching. Further investigations revealed that by using AdaBoost based feature selection, number of feature points used in classification were reduced by 75%.
Thus similar classification accuracy has been achieved at a significant reduction of computational cost.

Experiment 5: Matching of a selected set of features from training car images to that of test car image using the SIFT feature point matching approach proposed in Chapter 5.

A further experiment was performed to investigate the effects of SIFT feature selection when used in conjunction with the novel SIFT keypoint matching approach proposed in chapter 5.

From the results of experiment 3 and its analysis, it is obvious that the proposed AdaBoost based feature selection technique helps in pruning the outliers i.e. the least representative features and preserve the image’s uniqueness by keeping the most representative features. In novel matching scheme proposed in chapter 5, SIFT descriptors of all keypoints of every training image were compared with SIFT descriptors of all points, within a maximum likelihood area of the test images, centred at the point that corresponds location wise to the keypoint of the training image. The matching procedure thus involved matching of every keypoint from all training images of a class to that of the test image, making it computationally costly. Further, matching involved keypoints which may not represent the uniqueness of a given model. For example non-robust keypoints from a class that can be a result of noise or minor differences between two cars of the same make and model. Thus including the AdaBoost based feature selection step after feature extraction and before adopting the novel feature point matching scheme proposed in chapter 5, decreases the number of keypoints used for matching by 75%. Further an increased recognition accuracy of 92% is achieved.
6.4 CONCLUSION

This chapter proposed the use of adaptive boosting in selecting the most representative SIFT features of a given vehicle make and model for the effective use in vehicle MMR. It has been shown that the proposed selection of the most appropriate SIFT features allows a significant gain in the computational cost of previous SIFT based approaches to vehicle MMR. It has been shown further that the use of more relevant feature matching techniques allows significant relative gains in recognition accuracy when compared to the use of traditional SIFT keypoint matching algorithms. The algorithm has been tested on a publicly available database of car frontal views to enable easy comparison with existing and future vehicle MMR algorithms.

Though SIFT provides features which are invariant to scale, rotation and other variations, these features largely depend on distinctive regions such as blobs and well textured patches. In the application domain of vehicle MMR, cars are textureless and inter-class differences are very small. Using SIFT as the only reliable features for the recognition purposes may not provide the excellent discrimination capability. The need is to improve the description of an object through the use of some additional shape features. In this prospect, Contour like features play a significant role in representation of geometric shapes in images (particularly important for vehicle MMR). Supporting this idea, next chapter proposes a novel technique for feature extraction based on contours of images.
CHAPTER 7

Use of Contourlet Domain Localised Directional Feature Maps in Vehicle MMR

7.0 INTRODUCTION

This chapter presents a novel vehicle MMR algorithm that uses localised directional feature maps in Contourlet Transform[69] (see Chapter 3) domain. It is noted that Contourlet Transforms have been previously used for feature extraction in vehicle MMR by S.Rahati et.al. [33] (See Chapter 2). The originality of the contribution of this work is the use of, specific local texture features in the lowpass subband and the oriented image edge features across the scales of directional resolutions of Contourlet transformed image as features, rather than the direct use of Contourlet transform coefficients as features.

For clarity of presentation, this chapter is divided into several sections. Section 7.1 provides a conceptual comparison between multi-resolution image analysis techniques, i.e., Wavelet, Curvelet and Contourlet Transforms, which leads to a discussion on the motivation behind the use of Contourlet Transform domain feature extraction. Section 7.2 presents the proposed Contourlet based approach to vehicle MMR. Section 7.3 provides experimental results and a detailed analysis. Finally section 7.4 concludes with an insight into possible future improvements.

7.1 RESEARCH MOTIVATION

All vehicle MMR approaches proposed in the literature (see Chapter 2) are based on an initial stage of feature detection, where the detected features are subsequently used in matching and classification. Majority of the methods proposed in the literature use edge maps as features. However, the pixel domain edge extractors
used (e.g. Canny [108]) are limited in their performance and therefore fail in
accurately capturing the smooth curves/contours which are an important part of
most images. Kazemi et.al’s proposal of [32] which is a vehicle MMR method
based on Discrete Curvelet Transforms [61, 65, 66] (see Chapter 3) is the first
attempt to address this problem. Curvelet transforms provide a multi-resolution,
band pass and directional functional analysis method which are useful to represent
the image edges and curved singularities present in most natural images, more
efficiently. Therefore they are more accurate in representing curved edges as
compared to traditional wavelet transforms [66] (see Chapter 3). The multi-
resolution representation of edges and curved singularities enables effective feature
matching across the scales of the Curvelet transform, thus improving the robustness
of the vehicle MMR algorithm. However a major challenge in capturing the
geometry and the directionality in images comes from the discrete nature of the
data: the input is typically, sampled images defined on a rectangular grid. For
example, directions other than horizontal and vertical look very different on a
rectangular grid. In other words because of pixelization, the notion of smooth
contours on sampled images are not obvious. For this reason unfortunately, a
mathematical transform such as Curvelet transform that is initially developed in the
continuous domain and then later discretized for sampled data [69] is not effective
to be used with digital images.

Identifying the above shortcomings a new breed of transforms named Contourlet
transforms were proposed in [69] (see chapter 3) in the discrete form, as a simple
directional extension for wavelets. Contourlet transforms starts with a discrete-
domain construction and then studies its convergence to an expansion in the
continuous domain and is thus more suitable for digital image processing. It
provides improvements to 2-D separable wavelet transforms for representing images
with smooth contours in all directions (see Figure 7.1). Identifying this advantage,
in [33] the authors proposed the direct replacement of Curvelet transforms used in
[32] with Contourlet transforms in vehicle MMR. The idea was to pass each image
through a Contourlet filter bank converting the images to n-level multi-resolution
decomposition (see Figure 7.1). The feature detection and classification was subsequently performed on the decomposed images, rather than in the pixel domain.

![Figure 7.1: Multi-scale and multi-directional decomposition of a car image using contourlet transforms.](image)

In both of the approaches [32] and [33], features used are directly derived from the coefficients of sub-bands. Although an attempt has been made in exploiting the redundancy present in the Contourlet domain by using only pre selected sub-bands for image matching, the feature used in matching is the standard deviation of coefficient values of sub-bands. Unfortunately this measure ignores the localization of features that are an important property which often provides discrimination capability between classes in vehicle MMR. Therefore in this chapter a technique is proposed to use additional features of the Contourlet decomposition, namely, the local texture in the lowpass sub-band and the presence of oriented image edges across the scales of directional resolutions. Finally a Support Vector Machine (SVM) [82, 83] with a polynomial kernel is used as the feature classifier resulting in
an algorithm for vehicle MMR. [Note: a comparison of the performance of the proposed technique with that of [32] and [33] is presented in section 7.3].

7.2 PROPOSED APPROACH

Figure 7.2 illustrates a block diagram of the proposed approach to Contourlet based Vehicle type recognition. Details of the various modules are detailed subsequently:
CHAPTER 7

N (=300) Cropped and normalized, 128*128 pixel car images from test + training set

\( \hat{D} \) Contourlet feature maps obtained using the equations 1,2 and 3 given in section 6.2.3

\( \hat{L} \) directional bands and \( \hat{L}+1 \) equal-sized scale levels: one coarse level image and \( \hat{L} \) band-pass images.

SVM with Polynomial Kernel

Classification results

Figure 7.2: Proposed Recognition Algorithm

7.2.1 Pre-Processing of Test and Training Images

Initially all images were cropped using the procedure presented by V.Petrović and T.Cootes in [23], making sure that the cropped areas encapsulate all important
components of the frontal views, such as the grill, lights, badge, bumper area, etc., of the cars but eliminates background clutter as much as possible. Subsequently all cropped images are normalized in size to 128*128 pixels.

7.2.2 Contourlet Decomposition

In this stage, cropped images of cars are decomposed by using the Contourlet Transform. According to the theory of Contourlet Transforms [69], Contourlet decomposition is constructed by applying two successive decomposition stages. The first stage transforms the original image into a Laplacian Pyramid (LP) having $L+1$ scale levels. The second stage decomposes each LP scale level into $D$ directional sub-bands by using a directional filter bank. Inspired by the idea of extracting visually significant coefficients from [106,107], down sampling operation is discarded in the LP scheme. As a result $L+1$ equal-sized scale levels are obtained: one coarse level image and $L$ band-pass images. Same number of directions is obtained for each of $L$ scale levels, thus obtaining a total of $L \times D$ equal sized directional bands (see Figure 6.3). Experiments of the proposed technique use $L = 2$ and $D = 8$.

![Figure 7.3: Contourlet decomposition of an image of a car with equal sized scales and directions. The scale levels are represented by the colour blocks.](image)
CHAPTER 7

7.2.3 Features Extraction

In section 7.1, the importance of using a localized feature extractor as against a measure such as the variance of coefficients values of the sub-bands, was discussed. As a result within the research context of the proposed work a modified version of the localized feature extraction technique of M.Barni, F.Bartolini and A.Piva [107] has been adopted. For each frequency direction \( d \) of the contourlet decomposition, localized Contourlet features are extracted through the calculation of a directional map \( \text{Map}_d \) given by following equation:

\[
\text{Map}_d(i, j) = \text{Tex}(i, j)^\theta \text{Ed}(d, i, j)^\theta \tag{7.1}
\]

Where

\[
\text{Tex}(i, j) = \var{\hat{C}_{l+1}(4(i - 1) + 1 + y, 2j - 1 + x)}_{x = 0.1, y = 0.3} \tag{7.2}
\]

\[
\text{Ed}(d, i, j) = \frac{1}{2} \sum_{i = 0}^{1} \sum_{j = 0}^{1} [\hat{C}_{l+1}^2(i + y, j + x)]^2 \tag{7.3}
\]

In the above set of equations, \( \hat{C}_{l+1} \) represents the Contourlet sub-band at scale level \( \hat{l} \) and frequency direction \( \hat{d} \), while \( \hat{C}_{l+1} \) is the coarse level sub-image. 'Tex' and 'Ed' are respectively the local texture from the original filtered image (i.e. the coarse image) and the oriented edges across the scales of directional resolutions respectively. As a result of feature detection a total of \( D \) Contourlet feature maps are obtained (with localized properties), i.e. one for each direction. Note that the directional feature map \( M_d \) proposed in [107] consisted of a 'Brightness' term, but has been discarded in the above modified version. This is due to the reason that in
vehicle MMR, matching between images should be invariant to changes in brightness. Note that for experiments of this chapter, the values of $\alpha$ and $\beta$ were experimentally found out to be 0.1 and 0.2 respectively.

7.2.4 Dimensionality Reduction (Optional)

Though the feature matrix for each image, extracted as a result of the proposed technique, is much lower in dimension (128*128) than the traditional contourlet features obtained from multiple scales and directions, a further dimensionality reduction may optionally be applied by making use of the 2DLDA approach [54] presented in chapter 4. According to the theory of 2DLDA [54, 55] (see Chapter 3 and 4), it provides distinction between classes by decreasing inter-class variance and increasing between-class variance. Thus by preserving the features with high between-class variance and low inter-class variance, feature matrix for each image is further reduced to enhance the recognition speed of classifier. In section 6.3 the final classification accuracies obtainable with and without the use of 2DLDA on the Contourlet features extracted following the strategy described in section 6.2.3 will be presented.

7.2.5 Support Vector Machine Based Classification

Finally the feature maps obtained following the procedure described in sections 7.2.3 and 7.2.4 (optional) are used as the input to a Support Vector Machine (SVM) classifier first proposed by C.Cortes and V. Vapnik in [82] (see Chapter 3). A SVM is an effective method for general purpose pattern recognition and is a powerful classification tool.

The basic idea is to map input data into a high dimensional space and to find a separating hyperplane with a maximal margin. A support vector (SVM) kernel is
used to map the data from the input space to the high-dimensional feature space which facilitates the problem to be processed in a linear form.

In the literature [79, 82] it has been generally concluded that for most applications, a low degree polynomial kernel or a RBF kernel works quite well. In experiments performed for proposed technique, a polynomial kernel with degree 1 and 2 provided the best results. For theoretical details and mathematical derivations related to SVM, details are provided in [79, 82, 83].

7.3 EXPERIMENTAL RESULTS & ANALYSIS

**Data:** The experiments were conducted on a database of 300 images (frontal views) of cars belonging to 25 different classes [23]. Each class consisted of at least 11 images of different cars belonging to the same make and model. The images were cropped and normalized to 128*128 pixels.

**Experiment 1: n-fold cross validation** [88] for the selection of a kernel for SVM and value of ‘K’ for k-NN

The results of applying SVM, using the Polynomial and RBF as kernel parameters on the feature sets of section 7.2.3 and k-NN with various values of ‘k’ are illustrated in Table 7.1.

For SVM, it is concluded that the Polynomial Kernel of degree 1 provides the overall minimum average error across all n-fold cross validations (here n represents number of folds). Similarly for k-NN best results are achieved by setting $k = 2$. However, according to [95] (see Chapter 3), the effect of noise is more for small values of $k$. To reduce this effect $k = 5$ has been chosen for further experiments.
Table 7.1: Evaluation of SVM and k-NN performance against different kernels and value of 'k' using n-fold cross-validation (values of n are 25, 10, 6, 4, 2) respectively.

<table>
<thead>
<tr>
<th>Classifiers</th>
<th>25-fold cross validation</th>
<th>10-fold cross validation</th>
<th>6-fold cross validation</th>
<th>4-fold cross validation</th>
<th>2-fold cross validation</th>
</tr>
</thead>
<tbody>
<tr>
<td>SVM</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Polynomial kernel</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Degree 1</td>
<td>5</td>
<td>6</td>
<td>6</td>
<td>5</td>
</tr>
<tr>
<td></td>
<td>Polynomial kernel</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Degree 2</td>
<td>6</td>
<td>7</td>
<td>7</td>
<td>6</td>
</tr>
<tr>
<td></td>
<td>RBF Kernel</td>
<td>30</td>
<td>37</td>
<td>37</td>
<td>25</td>
</tr>
<tr>
<td>k-NN</td>
<td>K=2</td>
<td>12</td>
<td>11</td>
<td>12</td>
<td>15</td>
</tr>
<tr>
<td></td>
<td>K=5</td>
<td>13</td>
<td>11</td>
<td>12</td>
<td>17</td>
</tr>
<tr>
<td></td>
<td>K=10</td>
<td>15</td>
<td>15</td>
<td>16</td>
<td>19</td>
</tr>
<tr>
<td></td>
<td>K=25</td>
<td>23</td>
<td>21</td>
<td>26</td>
<td>28</td>
</tr>
</tbody>
</table>

Experiment 2: Effect of training size on classification accuracy

Initially an experiment was designed to determine whether increasing the size of the training set improves the classification accuracy. It was observed that when the training set size increases, the accuracy values improve. For example in Figure 7.4(c) when the training set size increases from 4 to 10, the average classification error reduces from approximately 20% to 10%.

Experiment 3: Comparison of the proposed technique with directly using contourlet coefficients

This experiment was designed to determine the effect of using the novel Contourlet feature selection criteria presented in section 7.2 and to compare it with directly using Contourlet coefficients from different combinations of scales-levels, as features. The results are presented in Figure 7.4. Comparing Figure 7.4 (g) with the rest of graphs illustrated in Figure 7.4 (i.e. (a) to (f)), one can observe that the
CHAPTER 7

proposed localized Curvelet feature selection criteria results in an improved classification accuracy when applied to vehicle MMR. An accuracy rate of around 94% is achieved as the training class size is improved to 10. The best results obtained when using the traditional approach is when all sub-bands are used (see Figure 7.4(f)). The highest accuracy level obtained is 90%, when the training class size is 10.

Automatic Approaches Towards Vehicle Make and Model Recognition 128
Experiment 4: Using feature representation and feature vector selection technique of F.Kazemi et.al [32] and S.Rahati et.al [33] on database of car images used for the proposed technique for comparison purposes

According to S.Rahati et.al [33], the best accuracy results are obtained when using scale level 3 and 4, divided into 8 and 16 directional sub-bands, respectively. Therefore a 24 element feature vector of standard deviations for each car image is obtained. The results are illustrated in Figure 7.5. Figure 7.5(a) shows the results
CHAPTER 7

when SVM ‘one against all’ [79, 82, 83] (see chapter 3) has been used, whereas Figure 7.5(b) uses SVM ‘one against one and all’.

![Graph](image)

**Figure 7.5 (a):** SVM ‘one against all’ classification error vs. training class for using standard deviation of scale-levels 3&4 of [33].

![Graph](image)

**Figure 7.5 (b):** SVM ‘one against one and all’ classification error vs. training class for using standard deviation of scale-levels 3&4 of [33].
CHAPTER 7

Comparing results of graph in Figure 7.5(a) and the results of the proposed technique illustrated in Figure 7.4(g), it is proved that the proposed algorithm outperforms [33] by achieving an accuracy figure of 94%, as against an accuracy figure of approximately 50%. Experimental results reveal that the use of SVM 'one against one and all' (see Figure 7.5(b)) gives the worse level of classification accuracy, i.e. approximately 40%.

Experiment 5: Application of 2DLDA on contourlet features extracted using the proposed technique

A further experiment was performed first to reduce the dimensionality of extracted features (see Section 7.2.3) using 2DLDA while enhancing the between-class variance and subsequently using the reduced feature set in classification (see Section 7.2.4). In Figure 6.6, the classification accuracies obtained after dropping 10 and 100 least eigenvectors (from a total of 128 eigenvectors) are compared with that obtained when 2DLDA is not used and used, but all eigenvectors are considered.

![Figure 7.6: Classification accuracy vs. training class size.](image)
The results illustrate that when a significant number of least eigenvectors are dropped (for example 100) the classification accuracy increases, particularly when the training class size is small (for example 2). The reduction in dimensionality has resulted in the use of the optimum discriminant features in classification, hence improving the accuracy. By dropping 100 least significant projection axes, a feature matrix of $128*128$ dimensions is obtained for each car image. This is substantially lower in dimension than the feature matrix obtained as a result of section 7.2.3, which is $128*128$. Hence a further advantage of reducing the dimensionality of the feature space is a significant reduction of the computational cost of classification.

The results illustrated in Figure 7.6 further indicate a relative increase in accuracy when using smaller training class sizes. This is due to the fact that smaller training class sizes result in an added difficulty of discrimination, which can be substantially improved by the use of 2DLDAs and a limited set of eigenvectors.

**Experiment 6: Comparison of the use of SVM vs. k-NN classifiers**

Figure 7.7(a) and 7.7(b) illustrate the accuracy results of using SVM and k-NN classifiers on feature sets extracted following procedures presented in sections 7.2.3 and 7.2.4, respectively.
The results of figures 7.7(a) and 7.7(b) clearly illustrate that SVM outperforms k-NN.
7.4 CONCLUSION

In this chapter a novel approach to automatic vehicle MMR has been proposed, based on a localized directional feature selection criterion, on the Contourlet transform domain. It has been shown that the use of localized features in Contourlet transform domain results in a substantial increase in classification accuracy as compared to using variance/standard deviation of Contourlet coefficient values. Further this increased accuracy is obtained at a significant reduction of computational cost.

It can be further concluded that 2DLDA can be used as a means for dimensionality reduction, without affecting the classification accuracy. This enables the reduction of the computational complexity of the classification stage substantially. Experimental results further concluded that SVM performs substantially better as a classifier in vehicle MMR as compared to k-NN ($k = 5$) classifier.
CHAPTER 8

Conclusions

8.0 INTRODUCTION

This chapter summarizes the key ideas presented in chapters 4, 5, 6 and 7, draws conclusions and emphasizes the important contributions made by the research presented in this thesis. It also gives an insight into possible future directions of research, particularly with the intention of further extending the functionality and efficiency of the proposed algorithms.

The main motivation to the research presented in the thesis comes from the observation that vehicle MMR is a relatively new research area and therefore an inadequate volume of research has been performed as yet. Through this thesis, solutions to some of the open research problems/challenges in vehicle MMR, particularly with reference to rotation, illumination, occlusion, scale variations and generally with reference to computational complexity, have been addressed. Novel techniques have been proposed which include local and global feature extraction/selection methods along with classification/matching schemes towards vehicle MMR resulting in significant rates of recognition accuracy. These techniques have been compared conceptually and/or experimentally with state-of-the-art techniques. The presentation of this thesis is organised such that each of the contributory chapters 4-7, is revisited as a separate section, with a brief overview highlighting the motivation, novelty and contribution of that work, an extensive analysis portraying the adequacies and limitations of the proposed technique and finally making specific conclusion related to the proposed approach.
8.1 CONCLUSION OF THE THESIS

The thesis presented four original contributions to the state-of-the-art in vehicle MMR. The first contribution presented in Chapter 4, is the novel application of 2DLDA [54, 55] to vehicle MMR. This approach has been previously successfully used in face recognition but the general differences between the datasets of face recognition and vehicle MMR domain required specific design considerations to be followed. The motivation behind use of 2DLDA was to project car images from pixel domain into a representation which can decrease within class variance and increase variance between different classes, an important factor for separating classes i.e. make/models from each other. In the past literature a detailed comparison between the effects of selectively using eigenvectors when using 2DLDA has not been investigated. This has lead to a lack of understanding about how to use Linear Discriminant Analysis effectively under varying illumination conditions and partial occlusion. It was shown that the best performance with the 2DLDA approach is obtained when the eigenvectors of lower significance are ignored in contrast to the improvement obtained when using a PCA based approach where the eigenvectors with the highest significance have to be ignored. For the test database of 200 car images of 25 different make-model classifications, a best accuracy of 91% was obtained with the 2DLDA approach. The best accuracy obtained when using a state-of-the-art PCA based approach was 85%.

Chapter 5 presented the second novel algorithm which was developed based on a thorough analysis of the use of traditional SIFT [58] feature extraction and matching strategy in vehicle MMR proposed by L.Dlagnekov [25]. The general motivation behind using SIFT [58] feature descriptors in vehicle MMR are their rotation, scale, occlusion and illumination invariance properties. However, experimental investigations revealed that the performance of SIFT matching as used in [25] degrades substantially with decrease in quality of training and test images. Therefore in the proposed work the matching strategy has been modified to be effectively used within vehicle MMR where database and test set consist of frontal
view cropped images of cars obtained from natural scenes. The proposed approach first identifies keypoints on all database images and attempts to find best matching point for each keypoint of each database image, within a windowed search area, i.e. a maximum likelihood area, of a given test image. Experimental results and in depth analysis was provided to justify the effectiveness of window based search for matching. It was shown that when the idea was tested on the database of 200 car images of 25 different make-model classification a 90% accuracy rate was obtainable, which is significantly more as compared to the accuracy rate obtainable via the use of traditional SIFT keypoint matching scheme adopted in [25].

Chapter 6 provided an extension to the work proposed in chapter 5 by introducing a features selection stage prior to matching. By using weight values attached to features obtained as a result of using the popular AdaBoost [97, 98] (as features selection technique) approach and entropy [113] as a measure for significance of these weights, distinctive features can be separated out from noisy/outliers. It has been shown that the proposed selection of the most appropriate SIFT features allows a significant gain in the computational cost of previous SIFT based approaches to vehicle MMR. It has been shown further that the use of more relevant feature matching techniques such as the novel approach presented in Chapter 5, allows significant relative gains in recognition accuracy when compared to the use of traditional SIFT keypoint matching algorithms [58]. The algorithm has been tested on a publically available database of car frontal views to enable easy comparison with existing and future vehicle MMR algorithms. Accuracy rates of up to 92% has been achieved when using the feature selection approach in conjunction with the matching approach of chapter 5, as compared to obtaining an accuracy rate of 82% when using the matching approach of [58].

Finally Chapter 7 presents the fourth novel algorithm which is an improvement to an approach proposed by S.Rahati et.al's [33] on contourlet transform [69] domain feature extraction. The motivation behind this work was to extract features suitable for representation of geometric shapes in images, i.e. those features that are
particularly important in vehicle MMR. Contourlet transforms provide multi-scale and multi-directional contour like features of images. Instead of directly using contourlet coefficients or variance of contourlet coefficients values [33] as features, which are non-robust features, the novel approach extracts local lowpass texture and multiscale directional edges as features. It was shown that the use of localized features in Contourlet transform domain results in a substantial increase in classification accuracy as compared to using variance/standard deviation of Contourlet coefficient values. Further, it was shown that the above increased recognition accuracy can be achieved at a reduced computational cost by reducing the dimensionality of the proposed features using 2DLDA.

8.2 FUTURE WORK

Although a number of novel contributions to the state-of-the-art in vehicle MMR have been proposed in this thesis it is possible to extend the work to further improve recognition rates and functionality of the algorithms.

All implementations of the proposed algorithms have been carried out using MATLAB. One desirable task is to implement the proposed systems in languages such as C or C++ that will increase speed of execution of the vehicle recognition tasks up to real-time rates. It is anticipated that the conversion of the MATLAB implementation to C/C++ has the ability to increase the speed eight fold.

In chapter 4, the performance of the proposed 2DLDA based approach to vehicle MMR has been analyzed under illumination variations and occlusions. Although the approach was able to handle significant amounts of illumination and occlusion variance, its performance under pose variation was limited. It is suggested that further research be carried out in identifying how rotation invariance can be introduced to this approach. One idea is to use a rotation invariant feature rather than individual pixel values, as features. Further integrating the proposed approach within a multi-classifier system can be considered.
In chapter 5, the search for a matching point to a SIFT keypoint of a test image was carried out within a fixed sized window area of maximum likelihood. However the size of the search window depends on the anticipated image rotation and/or translation present between the training and test images. Knowledge of the camera calibration information and image capture mechanism will allow the system to calculate bounds of the window size most appropriate to be used. This will lead to improvements in rotation/translation invariance and will increase the computational efficiency of the system. However, although the use of SIFT descriptors addresses the rotation and scale invariance problems, it is not fully invariant to illumination. In order to improve illumination invariance, a number of local feature descriptors such as edge, corner, shape and texture descriptors can be combined to the SIFT descriptor used above. This could also be improved by use of illumination/highlight removal algorithms as a preprocessing stage. However, steps should be taken to maintain rotation/scale invariance in light of the changes that are brought upon by other descriptors.

Humans have an ability to identify the make and models of vehicles based on identifying components, such as grilles, lights, badges and bumpers. Therefore component based identification of vehicle make and model provides a more natural approach. However an automated computer vision based approach to achieve component based vehicle MMR requires accurate segmentation of components which can be a challenge under certain lighting/illumination conditions and the presence of noise. However, if the test images are clear, such as that obtainable under controlled lighting conditions (e.g. within a multi-story car park with artificial lighting) accurate segmentation of components will be a possibility. Given that details of each component can be obtained from manufacturer's data such as high quality images, design data etc., accurate training of a component wise classifier will also be a possibility. Thus given that the above conditions are met a component based vehicle MMR technique can provide better rates of recognition. In chapter 6 the use of AdaBoosting in feature selection has been proposed. It was shown that...
CHAPTER 8

this approach provides a means to cluster distinct features of a particular vehicle make and model, component wise. A further improvement to this approach is possible via the use of more efficient interest point matching schemes.

The test image database used in this research consists of only the frontal view images and is comparatively small in size. Having multiple views allow the design and development of vehicle MMR techniques which are based on multiple angle views of a vehicle. Having a larger database with more samples per make-model will improve the accuracy of recognition considerably due to the possibility of improving the training process.

The fundamental theories of computer vision and its current present state of art provide a large number of feature detectors that can be combined with a large number of classifiers to provide the best recognition accuracy. It is recommended that research be carried out to explore the use of feature descriptors and classifiers not investigated in this thesis. To find the combination of feature descriptor/s and classifier pair that gives optimum accuracy remains a challenging and open research problem. Use of multiple descriptors in multi-classifier approaches is recommended.
REFERENCES


REFERENCES


REFERENCES


REFERENCES


REFERENCES


REFERENCES


REFERENCES


REFERENCES


REFERENCES


REFERENCES


REFERENCES


REFERENCES


REFERENCES


REFERENCES


APPENDIX

APPENDIX-A

Scholarly Contributions

The work presented in this thesis has resulted in a number of paper submissions/acceptances. The details can be given as follows.

REFEREED CONFERENCE PUBLICATIONS


A.4 I.Zafar, B.S.Acar, E.A.Edirisinghe, and H.E.Bez, Two Dimensional Statistical Linear Discriminant Analysis for Real-time Robust Vehicle Type Recognition,
APPENDIX


CONFERENCE PAPERS UNDER REVIEW


JOURNAL PAPERS TO BE SUBMITTED

