The thesis of Somesh Kashyap was reviewed and approved* by the following:

Yanxi Liu  
Associate Professor of Electrical Engineering and Computer Science and Engineering  
Thesis Adviser  

Robert Collins  
Associate Professor of Computer Science and Engineering  

David J. Miller  
Professor of Electrical Engineering  

W. Kenneth Jenkins  
Professor of Electrical Engineering  
Head of the Department of Electrical Engineering  

*Signatures are on file in the Graduate School.
Abstract

In this work we propose, implement and evaluate a discriminative computational pipeline to address the problem of feature generation, feature screening and feature subset selection from biomedical datasets when the ratio of feature dimensions (possibly in millions) to number of samples is very high. The proposed pipeline is modular and can be highly parallel. The framework is applied to a variety of real world discrimination problems including plant species classification from 2D leaf images, Gender/age/expression/human identification from 3D facial surface mesh and Gender/Age/Disease classification from 3D neural Magnetic Resonance Imaging (MRI) images. By using a unique set of novel features for each application we either achieve better or competitive classification accuracy than existing work or set new benchmarks otherwise. We can also locate discriminative regions for faces and brains which facilitate further scientific discoveries. This work illustrates quantitatively, the effectiveness and the diversity of the proposed feature extraction and machine learning pipelines.
Contents

List of Figures \hspace{1cm} x

List of Tables \hspace{1cm} xx

Acknowledgments \hspace{1cm} xxv

1 Introduction \hspace{1cm} 1

1.1 Motivation and Related work \hspace{1cm} 4

1.1.1 Overview of feature reduction techniques \hspace{1cm} 4

1.1.1.1 Feature Selection \hspace{1cm} 5

1.1.1.2 Feature Transformation \hspace{1cm} 6

1.1.2 Related work \hspace{1cm} 7

1.2 Proposed method \hspace{1cm} 10

1.3 Major contributions of the thesis \hspace{1cm} 11

1.4 Structure of the thesis \hspace{1cm} 14

2 Classification Framework \hspace{1cm} 16
2.1 Feature Normalization ................................................. 19
2.2 Feature Ranking ....................................................... 22
  2.2.1 Variance Ratio (VR) [64, 15, 26] ............................. 24
  2.2.2 Augmented Variance Ratio (AVR) [80, 88, 117, 120] .... 24
  2.2.3 Weighted Variance Ratio (WVR) ............................. 25
  2.2.4 Weighted Augmented Variance Ratio (WAVR) .......... 25
  2.2.5 Total Augmented Variance Ratio (TAVR) ................. 25
  2.2.6 Total Weighted Augmented Variance Ratio (TWAVR) . 26
  2.2.7 Weighted Ranking .............................................. 26
2.3 Redundancy Removal .................................................. 29
2.4 Features Subset Selection ........................................... 30
  2.4.1 Sequential Forward Selection (SFS) [56] .................... 31
  2.4.2 Sequential Backward Selection (SBS) [56] .................. 31
  2.4.3 Take L - Remove R (LR) [33] ................................. 31
  2.4.4 Sequential Floating Forward Selection (SFFS) [101] .... 32
2.5 Classification and Regression ....................................... 32

3 Learning from Leaves .................................................. 37
  3.1 Leaf Datasets ....................................................... 38
    3.1.1 Swedish leaf dataset ........................................ 38
    3.1.2 Ecotypes Dataset ............................................ 39
3.2 Related work ............................................. 40
  3.2.1 Choice of the Swedish leaf dataset ................. 43
3.3 Preprocessing ............................................. 45
3.4 Feature Extraction ....................................... 47
  3.4.1 Features from Shape .................................. 47
  3.4.2 Features from Texture ................................ 52
3.5 Results and Discussions ................................ 56
  3.5.1 Swedish Leaf Dataset ................................. 56
  3.5.2 Ecotypes Dataset ...................................... 58

4 3D Face Discrimination 63
  4.1 Related work ............................................. 64
  4.1.1 3D Gender recognition from face .................. 65
  4.1.2 3D Expression recognition from face ............... 65
  4.1.3 3D face recognition under expression variations ... 67
  4.1.4 Facial asymmetry ..................................... 69
  4.1.5 Shortcomings of existing approaches ............... 69
  4.2 3D Face Datasets ....................................... 71
   4.2.1 BJUT-3DR1 dataset (500 subjects) ............... 71
   4.2.2 German Dataset (200 subjects) .................... 71
   4.2.3 BU3DFE dataset (100 subjects) ................... 72
4.3 Preprocessing .................................................. 73
  4.3.1 Normalization ............................................. 74
4.4 Feature Extraction ........................................... 75
  4.4.1 Local Feature Extraction ................................. 76
    4.4.1.1 Local surface descriptors ....................... 76
    4.4.1.2 Asymmetry of Local surface descriptors ........ 78
    4.4.1.3 ICP residual error as a descriptor .......... 82
    4.4.1.4 Statistical of Local Features ................... 82
  4.4.2 Grid-Sampling and Multi-Scale representation .......... 83
  4.4.3 Global Feature Extraction ............................. 85
4.5 Results and Discussions .................................. 86
  4.5.1 Gender Recognition .................................... 86
    4.5.1.1 Comparison of Most Discriminating facial regions between male and female ....................... 93
    4.5.1.2 Importance of facial asymmetry .................. 93
  4.5.2 Age Estimation ......................................... 109
    4.5.2.1 Age Classification ............................... 109
    4.5.2.2 Age Regression ................................... 111
  4.5.3 Expression Classification .............................. 116
  4.5.4 Expression invariant gender classification .......... 120
  4.5.5 Face identification ................................... 124
# Neuroradiology Image Analysis

## 5.1 3D Structural brain MRI datasets
- **5.1.1 Aging Dataset**
- **5.1.2 Alzheimer’s Dataset (ADRC)**
- **5.1.3 Cardiovascular Health Study (CHS) Dataset**

## 5.2 Feature Extraction

## 5.3 Deformable Registration framework
- **5.3.1 Brain Atlas**
- **5.3.2 Brain Atlas Labels**
- **5.3.3 Preprocessing**
- **5.3.4 3D Deformable registration**

## 5.4 Feature Extraction

## 5.5 Results and Discussions
- **5.5.1 Gender Recognition**
- **5.5.2 Age Estimation**
- **5.5.3 Computer Aided Diagnosis (CAD) for Alzheimer’s Disease**
  - **5.5.3.1 ADRC**
  - **5.5.3.2 CHS**

## 6 Summary and Conclusion

## Appendix A. Moments

viii
List of Figures

1.1 Proposed Pipeline ............................................. 12
2.1 The classification pipeline ................................. 17
2.2 Comparison of feature normalization schemes ............. 22
2.3 Comparison of feature ranking schemes ...................... 27
2.4 Comparison of feature selection schemes ................... 33
2.5 Comparison of optimization criterion during feature selection .... 34
2.6 Comparison of various wrappers ............................. 35
2.7 Comparison of various classifiers ............................ 36
3.1 Sample leaves from the Swedish leaf dataset ............... 39
3.2 A Sample from the ecotypes dataset. The blue rectangle encloses the leaves of interest that we study. ......................... 41
3.3 Segmentation templates for leaf segmentation .............. 46
3.4 Preprocessing stages for ecotypes dataset ................. 46
3.5 Basic Leaf properties ......................................... 48
3.6 Curvature along Leaf boundary ............................. 51
4.10 Combined AVR and p-Value plot ........................................... 100
4.11 AVR vs p-Value plot ......................................................... 101
4.12 Age prediction on Test and Training set ............................. 114
4.13 Absolute Age prediction error on Test and Training set ........ 115
4.14 Feature Distribution in top 3 mRMR feature space during one random split for expression classification ......................... 119
4.15 Relative importance of local height features (101x101 grid) for various discrimination tasks. Note that features that are important for one task may not be for other discrimination task. This is similar to results reported by Liu et al.[81], where it was shown that different features were relatively important for face identification and expression recognition for the 2D case. Table 4.37 and 4.38 show some of the features mapped to the 3D face. ....................................................... 125
5.1 Age distribution of subjects in the aging dataset ....................... 129
5.2 Cross section of a subject from the aging dataset ..................... 130
5.3 Harvard Atlas .......................................................... 134
5.4 Preprocessing pipeline .................................................. 135
5.5 Cross section and 3D view of MNI152 brain that was used to register all the MRI scans in this dataset. The bottom row shows the labelled atlas. ................................................................. 136
5.6 Rendering of Brain Atlas ............................................. 137
5.27 Histogram of all feature stats for gender discrimination after SFS . . 154
5.28 Important regions for gender classification (local features) . . . . . 155
5.29 Important regions for gender classification (local features) . . . . . 156
5.30 Visualization of selected features for age classification . . . . . . 159
5.31 Age prediction on the Brain . . . . . . . . . . . . . . . . . . . . . . . . . . 160
5.32 Important regions for age classification (all features) . . . . . . 160
5.33 Age prediction on the Brain . . . . . . . . . . . . . . . . . . . . . . . . . . 161
5.34 Important regions for age classification (all features) . . . . . . 161
5.35 Age prediction on the Brain . . . . . . . . . . . . . . . . . . . . . . . . . . 162
5.36 Important regions for age classification (all features) . . . . . . 162
5.37 Age prediction on the Brain . . . . . . . . . . . . . . . . . . . . . . . . . . 163
5.38 Important regions for age classification (all features) . . . . . . 163
5.39 Age prediction on the Brain . . . . . . . . . . . . . . . . . . . . . . . . . . 164
5.40 Important regions for age classification (Asymmetry features) . . . 164
5.41 Age prediction on the Brain . . . . . . . . . . . . . . . . . . . . . . . . . . 165
5.42 Important regions for age classification (Asymmetry features) . . . 165
5.43 Age prediction on the Brain . . . . . . . . . . . . . . . . . . . . . . . . . . 166
5.44 Important regions for age classification (Local features) . . . . . . 166
5.45 Age prediction on the Brain . . . . . . . . . . . . . . . . . . . . . . . . . . 167
5.46 Important regions for age classification (Local features) . . . . . . 167
5.47 Histogram of all feature types for age discrimination after AVR ranking 168
5.48 Histogram of all feature types for age discrimination after SFS 168
5.49 Histogram of local feature types for age discrimination after AVR 169
5.50 Histogram of all feature types for age discrimination after SFS 169
5.51 Histogram of asymmetry feature types for age discrimination 170
5.52 Histogram of all feature types for age discrimination after SFS 170
5.53 Important regions for age classification (all features) 171
5.54 Histogram of all feature types for age discrimination after SFS 171
5.55 Important regions for age classification (local features) 172
5.56 Histogram of all feature types for age discrimination after SFS 172
5.57 Important regions for age classification (Asymmetry features) 173
5.58 Histogram of all feature types for age discrimination after SFS 173
5.59 Histogram of all feature stats for age discrimination after SFS 174
5.60 Histogram of all feature stats for age discrimination after SFS 174
5.61 Histogram of all feature stats for age discrimination after SFS 175
5.62 Histogram of all feature stats for age discrimination after SFS 175
5.63 Histogram of all feature stats for age discrimination after SFS 176
5.64 Histogram of all feature stats for age discrimination after SFS 176
5.65 Distribution of features in top 3 selected feature dimensions for ADRC dataset 178
5.66 Visualization of selected features for Alzheimer’s disease classification 179
5.67 Histogram of all feature types for alzheimer’s disease discrimination after AVR ranking ........................................ 180

5.68 Histogram of all feature types for alzheimer’s disease discrimination after SFS ranking ........................................ 180

5.69 Histogram of local feature types for alzheimer’s disease discrimination after AVR ranking ........................................ 181

5.70 Histogram of local feature types for alzheimer’s disease discrimination after SFS ranking ........................................ 182

5.71 Histogram of asymmetry feature types for alzheimer’s disease discrimination after AVR ranking ........................................ 183

5.72 Histogram of all asymmetry types for alzheimer’s disease discrimination after SFS ranking ........................................ 184

5.73 Histogram of anatomical regions selected for alzheimer’s disease discrimination after AVR ranking of all feature ........................................ 185

5.74 Histogram of anatomical regions selected for alzheimer’s disease discrimination after SFS ranking of all feature ........................................ 186

5.75 Histogram of anatomical regions selected for alzheimer’s disease discrimination after AVR ranking of local feature ........................................ 187

5.76 Histogram of anatomical regions selected for alzheimer’s disease discrimination after SFS local feature ........................................ 188
5.77 Histogram of anatomical regions selected for alzheimer’s disease discrimination after AVR ranking of asymmetry feature .......................... 189
5.78 Histogram of anatomical regions selected for alzheimer’s disease discrimination after SFS of local feature ......................................................... 190
5.79 Histogram of all feature stats for alzheimer’s disease discrimination after AVR ranking .................................................................................. 191
5.80 Histogram of all feature stats for alzheimer’s disease discrimination after SFS ranking .................................................................................. 191
5.81 Histogram of local feature stats for alzheimer’s disease discrimination after AVR ranking .................................................................................. 192
5.82 Histogram of local feature stats for alzheimer’s disease discrimination after SFS ranking .................................................................................. 193
5.83 Histogram of asymmetry feature stats for alzheimer’s disease discrimination after AVR ranking ........................................................................ 194
5.84 Histogram of asymmetry feature stats for alzheimer’s disease discrimination after SFS ranking ........................................................................ 195
5.85 Histogram of all feature types for CHS1 disease discrimination after AVR ranking .................................................................................. 198
5.86 Histogram of all feature types for CHS1 disease discrimination after SFS ranking .................................................................................. 198
5.87 Histogram of anatomical regions selected for CHS1 disease discrimination after AVR ranking of all feature .......................... 199
5.88 Histogram of anatomical regions selected for CHS1 disease discrimination after SFS ranking of all feature .......................... 200
5.89 Histogram of all feature stats for CHS1 disease discrimination after AVR ranking ............................................................... 201
5.90 Histogram of all feature stats for CHS1 disease discrimination after SFS ranking ............................................................... 201
5.91 Histogram of all feature types for CHS2 disease discrimination after AVR ranking ............................................................... 202
5.92 Histogram of all feature types for CHS2 disease discrimination after SFS ranking ............................................................... 202
5.93 Histogram of anatomical regions selected for CHS2 disease discrimination after AVR ranking of all feature ....................... 203
5.94 Histogram of anatomical regions selected for CHS2 disease discrimination after SFS ranking of all feature ....................... 204
5.95 Histogram of all feature stats for CHS2 disease discrimination after AVR ranking ............................................................... 205
5.96 Histogram of all feature stats for CHS2 disease discrimination after SFS ranking ............................................................... 205
5.97 Histogram of all feature types for CHS3 disease discrimination after AVR ranking ........................................ 206
5.98 Histogram of all feature types for CHS3 disease discrimination after SFS ranking ........................................ 206
5.99 Histogram of anatomical regions selected for CHS3 disease discrimination after AVR ranking of all feature .................. 207
5.100 Histogram of anatomical regions selected for CHS3 disease discrimination after SFS ranking of all feature .................. 207
5.101 Histogram of all feature stats for CHS3 disease discrimination after AVR ranking ........................................ 208
5.102 Histogram of all feature stats for CHS3 disease discrimination after SFS ranking ........................................ 208
6.1 Screenshot of pipeline output ........................................ 212
List of Tables

2.1 Selected classification results .................................................. 18
2.2 Independent feature ranking schemes ......................................... 24
2.3 Discriminating regions based on different ranking schemes .......... 28
2.4 Comparison of different classifiers as wrappers for feature selection.
Each row indicates the classifier that was used for feature selection,
while each column indicates the final classification rate using different
classifiers. The best classification in each row has been highlighted.
$X$-component of the deformation field, as described in section 5.4 of Chapter 5 was used for this experiment. After AVR feature ranking, mRMR
was used to select top 100 independent features, followed by SFS. The
best result was achieved when LDA was used as a wrapper as well as
the final classifier. NOTE: MAH = Mahalanobis, KNN= K-nearest
neighbor, LDA = Linear Discriminant Analysis, QDA = Quadratic
Discriminant Analysis, SVM = Support Vector Machine (linear kernel) 36

3.1 Previous work Plant Specie classification ................................. 44
3.2 Statistical moment features ........................................ 50
3.3 Swedish dataset classification results ............................. 57
3.4 Top AVR features ....................................................... 59
3.5 Ecotypes classification results ....................................... 59
4.1 Recent work on 3D discrimination ................................. 68
4.2 Differential geometry based surface features ..................... 79
4.3 Differential geometry based surface features (cont...) .......... 80
4.4 Statistical Moments for face discrimination .................... 83
4.5 Datasets for 3D face classification experiments ................. 87
4.6 3D facial gender classification results ............................ 88
4.7 List of symbols for 3D facial surface features Note that Orientation (O)
     , Direction of maximum curvature U(X) and Direction of minimum
     curvature U(N) have X, Y and Z components. .................... 89
4.8 Selected Features for gender discrimination ..................... 90
4.9 Selected Features for gender discrimination ..................... 91
4.10 Selected Features for gender discrimination .................... 92
4.11 Discriminating regions for gender classification (1) .......... 94
4.12 Discriminating regions for gender classification (2) .......... 95
4.13 Discriminating regions ............................................. 96
4.14 Discriminating regions ............................................. 97
4.15 Discriminating regions ............................................. 98
4.16 p-Value scores showing statistically significant regions for gender discrimination corresponding to Table 4.14. p-Value above 0.05 has been made transparent, while p-Value below 0.001 has been saturated to 0.001. Only Height(Z), Orientation Y component (Oy), Mean Curvature (H), absolute value of mean curvature (abs(H)) and Shape Index (S) are shown here. 99

4.17 p-Value plot of most discriminating regions 103
4.18 p-Value plot of most discriminating regions 104
4.19 Histogram of mean features 105
4.20 Histogram of mean features 106
4.21 Histogram of mean features 107
4.22 Histogram of mean features 108
4.23 Results for age classification 109
4.24 Discriminating regions for Age 110
4.25 Discriminating regions for Age 112
4.26 Discriminating regions for Age 113
4.27 Confusion matrix showing expression classification rates and standard deviation averaged over 100 iterations on the BU3DFE dataset. The average overall classification rates is 88.87±3.15%. The best classification rate reported on this dataset is by Tang and Huang [116], who achieved 95.1% by using features extracted from 83 manually labelled fiducial points.

4.28 Level 3 Confusion matrix showing mean expression classification rates averaged over 10 iterations on the BU3DFE dataset. The average overall classification rates is 78.49%.

4.29 Level 2 Confusion matrix showing mean expression classification rates averaged over 10 iterations on the BU3DFE dataset. The average overall classification rates is 73.59%.

4.30 Level 1 Confusion matrix showing mean expression classification rates averaged over 10 iterations on the BU3DFE dataset. The average overall classification rates is 67.71%.

4.31 Confusion matrix for the classification rates (over all 84.0286±4.7%) when only full face features were used.

4.32 Confusion matrix for the classification rates (over all 79.03±5.8%) when only asymmetry features were used.

4.33 Gender classification rates under different expressions.

4.34 Discriminative regions for gender identification (local features).
4.35 Discriminative regions for gender identification (asymmetry features) 123

4.36 Classification rates ± standard deviation averaged over 100 iterations when using (a)symmetry features alone, local features alone and when both are combined. 124

4.37 Relative importance of local Features 126

4.38 Relative importance of asymmetry Features 127

5.1 Brain gender classification results 144

5.2 Brain age regression results 158

5.3 Classification results on ADRC dataset 177

5.4 Number of subjects in CHS datasets 196

5.5 Classification results on CHS 1 dataset 196

5.6 Classification results on CHS 2 dataset 196

5.7 Classification results on CHS 3 dataset 197

5.8 Classification results on CHS datasets 197

5.9 Classification results on CHS datasets (balanced) 197

6.1 Selected classification results 213
Acknowledgments

I would like to thank Prof Yanxi Liu for her constant support and suggestions without which this thesis would not have been possible.
Chapter 1

Introduction

‘The goal is to transform data into information, and information into insight’

- Carly Fiorina, Former HP chairman and CEO [32]

The elements of scientific process are observation, hypothesis formulation, prediction and validation [133]. Traditionally the “observation” in this process was limited by the human eye. However over the last few centuries scientists have extended our observation capabilities to see objects ranging from small atoms to large galaxies. This has also led to exponential growth in data to be analyzed [45, 87, 24, 27, 48, 122, 108]. A number of methods have been proposed for automatic understanding of the images [39, 130, 134, 22, 3, 124]. These methods extract a variety of features from the images and build models based on these features. However, most of these methods fail as the number of features set size increases exponentially compared to number of training samples.
In this work, we propose to construct a five stage supervised learning pipeline [34, 18, 26] for analysis of biomedical datasets. The pipeline can be used for discriminating binary or multi-classes. If the classes are discrete variables, a classification-based approach is used otherwise a regression-based model is constructed. A predictive model can be validated on unseen data through cross-validation [58, 55, 103]. Cross-validation involves partitioning the dataset into multiple subsets. The first subset, called the training set, is used to build the predictive model while the other subsets, called the test/validation set, are used for validating the model. To reduce variability the process is repeated multiple times with different partitions. The advantage of a supervised learning system is that no prior knowledge is required to build a predictive model as the algorithm learns directly from the training data. If some domain specific prior knowledge is available, it can also be incorporated to improve accuracy of the model. Such algorithms have also made possible to identify patterns hidden in data which were impossible to find by a human observer, for example identifying regions of the brain for gender or age discrimination [83, 62]. The input to such an algorithm is a set of features extracted from training data. Prior knowledge is usually incorporated in the feature extraction process. However in most cases, it is not known apriori which feature might be useful for a particular discrimination task. Thus a large variety of features are extracted and input to the algorithm. The output is a predictive model based on a subset of discriminative features only. A minor variation to this approach is when we already have a feature list and we want
to identify whether a feature is useful for a discrimination task or not, e.g. identifying the genotype (gene combinations) which affects a phenotype (physical traits). In either case the initial feature set size can become exponentially large, especially with higher dimension images (e.g. MRI/f-MRI, Gene Array etc) [54]. The problem is further complicated if the number of samples is much less than the feature set size. The “true” dimension of the features is much less than the initial feature set size in practice, as many of the features are highly co-related and thus redundant [141]. Also many of the features are not discriminative. However these are not known beforehand. Thus the entire feature space has to be searched to find the optimal discriminative feature subset. A large training set is needed to search for an optimal feature subset in the N-dimensional hyper-volume. For a large number of machine learning problems, the number of samples is much lower than feature dimensions. Thus a small number of samples are used to estimate the true distribution of the features. This is a case of “The curse of dimensionality” [9, 13], which in general refers to the exponential increase in computing a task of interest when extra dimensions are added to an associated mathematical space. The term was first coined by Bellman [9]. To address this problem, we need to select a subset of features without significantly degrading (or even improving) the classifier performance [13]. As exhaustively searching for this subset is computationally prohibitive, selecting relevant features heuristically becomes a critical step in the machine learning process. In this work, we propose a classification pipeline to address this problem. The next section
gives a short overview of existing feature selection techniques followed by recent work which addresses the problem of feature selection from huge feature sets.

1.1 Motivation and Related work

Over the last decade, a lot of research effort has been put to improve feature selection techniques [41, 71, 72, 85, 73, 50, 101]. For an in-depth survey of feature selection techniques, the reader is referred to [105, 70]. Most of the feature subset selection methods evaluate subsets of features based on some criterion and output a subset which optimizes this criterion. While such methods work for feature set size of the order of hundreds, they do not scale to higher feature set sizes due to exponentially large number of feature combinations that needs to be evaluated. Thus prior to subset selection, the feature set size is often reduced by some ranking scheme.

1.1.1 Overview of feature reduction techniques

Feature reduction techniques are classified based on whether the feature mapping is preserved or not after the reduction. Feature selection [105] techniques preserve original feature mapping while feature extraction [42] technique transform the feature set into new set of features.
1.1.1.1 Feature Selection

Feature selection techniques can be divided into three broad categories as listed below.

**Filter methods** [95] Filter methods rank features based on some criterion independent of a classifier. These can further be divided into univariate and multivariate methods. Univariate methods rank individual features based on some criterion such as Chi-square, p-Value, information gain etc. Such methods are highly scalable but ignore feature dependencies. They are preferred if the input feature dimension is huge or if the individual feature mapping has to be maintained. However, as these methods do not take feature dependencies into account, a large number of output features are co-related thus highly redundant. There is also a lack of robust univariate ranking schemes for multi-class problems. Multi-variate methods such as Correlation based feature selection (CFS) [43], Markov blanket filter (MBF) [60] take feature dependencies into account but are not scalable as univariate methods.

**Wrapper methods** [59] Wrapper based methods include the classifier in the decision process. Features (single or a subset) are added/removed to a set if classification rate improves when considered with already selected feature subset. These methods can further be divided into deterministic (SFS [56], SBS [56], SFFS [101], Take l- remove R [33] etc.) and randomized (stimulated annealing). Deterministic methods are more prone to be stuck in local optimum
while randomized algorithms are less prone. Both methods have high risk of over training, especially if the number of feature dimensions is much larger than the number of training samples. They are also not scalable due to their combinatorial complexity.

**Embedded methods** [105] Such methods use a simple classifier (such as decision trees) and infer important variables from learner structures. For stability such methods often use ensemble methods (e.g. random forests). While scalable to some degree, such methods are more complex for multi-class and regression problems.

### 1.1.1.2 Feature Transformation

Feature transformation is another popular way for reducing the number of features. Feature transformation algorithms transform the feature set by applying a mapping from high dimensions to lower dimensions. Thus, the resulting feature set is a combination for all features, e.g a linear transformation via Principal Component Analysis (PCA) [86]. These methods have the following disadvantages:

- Individual feature information is lost. For example if a feature corresponds to a physical location on the face, then doing PCA on the feature set results
in loss of this spatial information.

- The computation time is not linear with feature dimensions
- The process involves working on the whole feature dimension, thus it is often difficult to make such algorithms parallel
- They have huge memory and processing requirements.
- Often the feature matrix is not full rank, hence resulting in numerical instabilities.

While each of these methods have their advantages and disadvantages, there is a need to a systematically combine each of these into a unified framework. The next subsection review some of the recent work which focus on feature selection from high-dimensional feature set size.

1.1.2 Related work

In the field of feature selection, the perception of “how large is large enough” has constantly been increasing over the years. This perception would continue to be revised, especially for high-dimensional image data, where the number of features that can be extracted is practically limitless. In 2000, Brunzell et al [17] applied a feature selection method based on an upper bound on the bayes classification error for a dataset with large number of features but few training samples. The highest dimensions they experimented with was 56. Jose et
al[12, 51] (2001) proposed a 3-stage pipeline for handling large feature dimensions for computer vision based systems. In the first stage they used Relief algorithm to remove irrelevance, K-means clustering was used in the second stage to remove redundancy followed by a standard combinatorial feature selection stage in the last stage. They showed this 3-step combination to be more effective than standard feature selection algorithms for huge feature dimensions. Working with a maximum feature dimension of 4096, they showed that the feature set could be reduced to 5% of its original size, with very little loss of information. The first work on large (4000) feature subset selection for AD and Schizophrenia MR image classification was done by Liu et al[82]. Palmer and Liu [80] (2003) used a combination of Augmented Variance Ratio(AVR) and Sequential Forward Selection (SFS) on a 3D face dataset of dimension 15686. Rangarajan et al[102] (2004) proposed a combination of regression lines and clustering for dimensional reduction of temporal data. Liu [75] (2005) proposed a embedded method called Basic Sort-Merge Tree (BSMT) for feature selection and applied it for video retrieval. Hong and Cho [46] (2006) used a modified representation of the chromosome in genetic algorithms and applied it cancer classification from DNA microarray data. Their definition of “huge feature dimension” was greater than 1000 features. Le and Satoh [63] (2006) proposed a filter method based on Conditional Mutual Information (CMI) for feature selection. CMI measure was shown to maximize relevance and minimize mutual
dependency on a feature set of dimension 134736. Zhang et al [143] proposed a combination of mRMR and Relief for gene array analysis and applied it to a feature set of dimension 12582. The highest feature dimension used so far has been by Liu et al [83] (2007), where they investigated a feature dimension of over 500 million on a sample size of 66 subjects. They used AVR for feature screening followed by SFS. Teverovsky et al [120] (2008) used 11,479,470 brain asymmetry features for predicting age from brain MRI. They proposed a combination of AVR ranking and forward feature selection. Relief was used for irrelevance removal while mRMR was used for redundancy removal. Liang et al [66] (2008) proposed a Feature Selection based on Distance Discriminant (FSDD). The basis of discriminant function which increases inter class separability while minimizing intra-class distance between features. FSDD They proved that FSDD was able to find near optimal feature subset without exhaustive search or branch and bound. Peng et al [100] (2009) proposed a combination of filter and wrapper based approach for feature selection. Receiver Operating Characteristic (ROC) characteristics was used for feature ranking followed by Sequential Floating Forward Selection (SFFS) on the reduced feature subset. However they demonstrated the proposed approach only on a feature set of dimension 44 or less. Habibi and Saeid [4] (2009) demonstrated a stimulated annealing based feature selection on a feature set of size around 14 million. Liu et al [74] (2009) proposed a genetic selection method called Recursive Feature Ad-
dition, combining supervised learning and statistical similarity measures. Singh et al\cite{110}(2009) proposed a Single Feature Optimization based heuristic to add a feature to a subset and showed that the proposed method can be made highly parallel, hence scalable to large datasets. Gheyas and Smith \cite{35} (2010) proposed a new algorithm SAGA by combining stimulated annealing and genetic algorithm and demonstrated their method on a feature dimension of 10000.

Our proposed pipeline addresses some of the drawbacks of existing methods as described above. Lack of publicly or commercially available softwares which can handle such huge feature dimensions is a strong motivation for this thesis.

1.2 Proposed method

The proposed method (Figure 1.1) is systematic extension of work done by Liu et al\cite{83}. We address the problem of feature generation, feature screening and feature subset selection from biomedical datasets when the ratio of feature dimensions (possibly in millions) to number of samples is very high. Figure 1.1 gives an overview of the proposed pipeline.

I Feature Extraction The input to the pipeline is a initial set of features.

The feature extraction is not generic and depends of the dataset.
II  **Feature Scaling** The features are first scaled to prevent any feature with high dynamic range to dominate over that having low dynamic range.

III  **Univariate Feature Ranking** After scaling, the features are ranked individually based on some discriminative criterion.

IV  **Redundancy Removal** The output of the feature ranking block consists of many features which are co-related. The redundancy removal block eliminates features which are similar.

V  **Wrapper based feature selection** A wrapper based feature subset algorithm is then applied to select the final subset.

VI  **Predictive Model** The output of the pipeline is a predictive model which is evaluated on the validation set.

The proposed pipeline is modular and can be highly parallel. The pipeline can be used for discriminating binary or multi-classes. Chapter 2 describes the elements of the pipeline in more details.

### 1.3 Major contributions of the thesis

**Classification pipeline** The idea of novel feature extraction, feature ranking, feature subset selection followed by discrimination via indexing/ classification/ regression has been proposed and implemented before [81, 80, 88,
Figure 1.1: The block diagram of proposed pipeline is shown. The input to the pipeline is a feature set (I). The feature extraction process depends on the image modality. The output of the pipeline (VI) is a predictive model which can be tested on a validation set. The shaded blocks are the main part of the pipeline. These blocks are independent of the image modality. Section 1.2 explains these blocks briefly. For a detailed explanation, refer to Chapter 2.

The specific contribution of this thesis is to make this process systematic, modular and parallel. By systematically combining filter and wrapper based feature selection technique, we propose a classification pipeline capable of handling very large feature dimensions. We also include a redundancy removal stage. Besides the proposed pipeline, the novelty of the thesis also lies in the applications. We choose three different biomedical data types with successively increasing feature set size to show the power of the proposed pipeline.

**Plant leaf analysis**  We propose features based on leaf shape and texture (more than 3,000 features) for plant leaf analysis and use those features for automatic plant specie classification. The feature ranking stage of
the pipeline also identifies which features are more useful in discriminating different species. We apply our pipeline on two ecotypes of the same species, one grown in controlled environment and other wild ecotype. By this experiment we identify which features are controlled by environment and which ones genetically. Such a study is extremely useful for biologists whose interest lie in liking a particular phenotype to a genotype, e.g. which gene control the leaf shape.

**3D Face discrimination** We propose features based on surface height, orientation and curvature (above 300,000) for 3D facial analysis and apply it to the problems of gender, age, expression and person identification. We do an in-depth study of inter/intra race gender differences and identify regions of the face which differ between the groups. We have achieved the best classification rate reported so far in literature for 3D facial gender classification and age predication from 3D face. We have achieve competitive classification rates for expression, age and person identification from 3D mesh. We also identify which features are important for various face discrimination tasks. Identifying regions of the face which discriminate gender, age, race etc is of particular interest to many psychologists and anthropologists.

**3D Brain analysis** We build on the work done in [119, 120, 83] to extract many novel features from the brain MR images. We use features extracted
from the deformation field and the intensity (above 90 million) to many
novel applications such as gender and age estimation from brain. We fur-
ther identify which regions of the brain are more important for a particular
discrimination task. This provides a better understanding of the human
brain. We also apply our framework for identifying regions of the brain
which are affected by Alzheimer’s disease and develop a framework which
could be used for computer-aided-diagnosis of neorogenerative diseases.

**Importance of Asymmetry** Following up on the work of [81, 79, 80, 88, 117,
120, 83, 119] on quantified face and brain asymmetry and using an im-
proved Mid-Sagittal Plane (MSP) extraction algorithm [121] , we invest-
tigate the importance of asymmetry features for 3D face and volumetric
brain datasets.

## 1.4 Structure of the thesis

The rest of the thesis is organized as follows. Chapter 2 describes the modules of
the proposed pipeline in more details. Chapter 3, 4 and 5 show the application of
the pipeline for discrimination of Plant, 3D face and Brain images respectively.
Domain specific feature extraction, related work and results can be found in
these three chapters. Finally, Chapter 6 concludes the thesis and discusses the
future scope of this work. Appendix on moments can also be found at end of this
document. Due to the large volume of results, we present only the major and relevant results in this document. Readers are referred to the supplementary material for exhaustive results.
Chapter 2

Classification Framework

Supervised machine learning is a scientific discipline which involves development of algorithms that learn patterns from training data. Formally, given a set ,

\[ X = \{ x_1, x_2, \ldots, x_n \} \]

and corresponding labels,

\[ Y = \{ y_1, y_2, \ldots, y_n \} \]

a supervised machine learning algorithm maps from \( X \) to \( Y \), \( f(X) = y \). If the output \( y \) is continuous, the process is termed “Regression”, otherwise it's called “Classification”. The key steps of learning process are:

1. Control data collection
2. Novel feature extraction
3. Selecting discriminating features
4. Choosing the classifier
5. Testing the classifier
Figure 2.1: The classification pipeline

While step 1, 2, 4 and 5 have been researched extensively [49, 26, 13], it is only in the last decade that researchers have focused their efforts on step 3. One of the reasons for this is the emergence of high dimensional datasets such as MR/CT images, Gene Micro-Array etc. in recent years. Chapter 1 discussed feature selection and associated problems briefly. We also briefly described the proposed pipeline to handle large feature set size. This Chapter discusses the individual modules of proposed pipeline in greater detail.

Figure 2.1 shows the block level view of the proposed pipeline. The input to the pipeline is a set of domain specific features and the output is a predictive model. In this chapter we describe the generic classification pipeline. Domain specific feature extraction is described in chapter 3, 4 and 5 for plants, 3D face and Brain respectively.

Chapter 1 discussed some of the problems associated with feature selection in high dimensional space. Table 3.1 lists the approximate initial feature set size
<table>
<thead>
<tr>
<th>Domain</th>
<th>Image Type</th>
<th>IFS</th>
<th>FFS</th>
<th>Discriminant</th>
<th>Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>Plant specie</td>
<td>2D color image</td>
<td>$10^4$</td>
<td>20-30</td>
<td>LDA</td>
<td>98.93($\pm 2.3\text{%}$)</td>
</tr>
<tr>
<td>Face Gender</td>
<td>3D surface mesh</td>
<td>$10^6$</td>
<td>40-60</td>
<td>LDA</td>
<td>96.3($\pm 3.5\text{%}$)</td>
</tr>
<tr>
<td>Age Classification from face</td>
<td>3D surface mesh</td>
<td>$10^6$</td>
<td>20-30</td>
<td>LDA</td>
<td>88.8($\pm 5.7\text{%}$)</td>
</tr>
<tr>
<td>Age Regression from face</td>
<td>3D surface mesh</td>
<td>$10^6$</td>
<td>40-60</td>
<td>Linear</td>
<td>4.0($\pm 3.4\text{ years}$)</td>
</tr>
<tr>
<td>Face Expression</td>
<td>3D surface mesh</td>
<td>$10^6$</td>
<td>50-90</td>
<td>LDA</td>
<td>88.87($\pm 3.1\text{%}$)</td>
</tr>
<tr>
<td>Brain Gender</td>
<td>Structural volumetric MRI</td>
<td>$10^8$</td>
<td>40-60</td>
<td>LDA</td>
<td>87.8($\pm 4.5\text{%}$)</td>
</tr>
<tr>
<td>Brain Age</td>
<td>Structural volumetric MRI</td>
<td>$10^8$</td>
<td>40-60</td>
<td>Linear</td>
<td>3.8($\pm 3.0\text{ years}$)</td>
</tr>
<tr>
<td>ADRC</td>
<td>Structural volumetric MRI</td>
<td>$10^8$</td>
<td>10-25</td>
<td>LDA</td>
<td>87.4($\pm 15.0\text{%}$)</td>
</tr>
<tr>
<td>CHS (average)</td>
<td>Structural volumetric MRI</td>
<td>$10^8$</td>
<td>5-20</td>
<td>LDA</td>
<td>64.6($\pm 40.5\text{%}$)</td>
</tr>
</tbody>
</table>

Table 2.1: Input image type, initial feature set size, average selected feature set size and selected classification results. For all the experiments, AVR was used for univariate ranking, followed by mRMR and SFS feature selection with a LDA wrapper. Note IFS= Initial feature set size, FFS= Final feature set size.

and the size of the finally selected features for some of the datasets we use in this work. It can be seen that by only using a sparse set of 40-70 features selected by the proposed pipeline, high classification accuracy is achieved on all datasets.

As can be seen from Table 3.1, the initial feature set size can be huge ($> 10^6$). This huge set of features limits the possibility of using all features together due to memory and processing limitations. Not all the features are equally discriminative, thus inclusion of non-discriminative features would certainly not
improve, if not degrade, the classification accuracy, besides imposing a huge burden on memory and processing requirements. Thus, we need to select a suitable subset of the features for the classification task at hand. Searching exhaustively for this subset is computationally prohibitive, hence we propose a 4-stage classification pipeline as shown in Figure 2.1.

The four modules of the pipeline are:

1. Feature normalization
2. Feature ranking
3. Feature redundancy removal
4. Wrapper Feature subset selection

Each stage is described in detail below.

2.1 Feature Normalization

Feature normalization refers to independent scaling of features. It has been shown by [25] that feature scaling improves the classification rates for SVM. Similar considerations apply to other classifiers as well. The main advantage of feature scaling is that it prevents features with high dynamic range (e.g. range data from 3D mesh) from dominating over features having low dynamic range (e.g. 3D facial mesh curvature). It also prevents numerical instabilities.
The scaling parameters are calculated from the training set and the same scale is used for the test set. Thus if we scale the training features to \([0,1]\), some features in the test set may have values -0.1 or 1.2. The proposed pipeline has six choices for feature scaling:

**No scaling** If all features have the same dynamic range (and the range is not very large), then scaling is not necessary. For instance, when using only texture features for a classification task, all the features would lie in the interval \([0,255]\), hence feature scaling can be skipped.

**Log scaling** If the range of feature values is too low or too high, the features can be scaled to have a narrower range by taking log of the features.

**Fixed range scaling** In this scheme all the features are scaled to a fixed interval \([a,b]\) chosen by the user. Thus the maximum value of a feature maps to “\(b\)” while minimum value maps to “\(a\)”. All other features are mapped in between. A possible downside of this ranking scheme is that if one sample in the training set has unusually high value, then all other features would be compressed to a very small dynamic range. Popular choices for the range \([a,b]\) are \([0,1]\) and \([-1,1]\).

**Gaussian scaling** In this ranking scheme the mean of the feature is sub-
tracted from the feature and it is divided by the standard deviation.

\[
x = \frac{x - \text{mean}(x)}{\text{std}(x)} \quad (2.1)
\]

The mean may be replaced by median if few samples are believed to be noisy.

**Min-Max scaling** In this ranking scheme the feature is transformed according to equation 2.2.

\[
x = \frac{x - \text{min}(x)}{\text{max}(x) - \text{min}(x)} \quad (2.2)
\]

**S-scaling (softmax)** Softmax scaling is a non-linear scaling and has a scaling factor in the shape of “S”, reaching both minimum and maximum values asymptotically. While the values in the middle are almost linearly transformed, the values in the ends are squashed into a narrow range. A logistic function given by 2.3 can be used to generate a “S” curve. The feature transformation is given by equation 2.4.

\[
L(x) = \frac{1}{1 + e^{-ax}} \quad (2.3)
\]
Figure 2.2: The classification rates averaged over 100 iterations when various feature normalization schemes were used. X-component of the deformation field, as described in section 5.4 of Chapter 5 was used for this experiment. After feature ranking, mRMR was used to select top 100 independent features, followed by SFS. LDA was used as the final classifier. It can be seen that Zero-to-One scaling performed best.

\[
x = \frac{1}{1 + e^{-\frac{\text{mean}(x) - x}{\text{std}(x)}}}
\]

(2.4)

2.2 Feature Ranking

Traditional dimension reduction techniques such as Principal Component Analysis (PCA) \cite{86}, Local Linear Embedding (LLE) \cite{107}, etc. cannot be used on feature dimensions of the order of 90 million due to processing and memory requirements. Also the complete feature matrix is not usually full rank and
invertible, making traditional dimension reduction techniques of limited use. Thus to reduce the dimensionality of the features as well as to weed out non-discriminative features, we rank the features based on different discriminative criterion and select the top n% of features, “n” is specified by the user. Such a ranking doesn’t transform the original feature set, but only selects features individually. Such a requirement is often important for applications such as genetic-analysis where maintaining the identity of genes is crucial. In the context of our work, it helps to map a feature back to its spatial location. Table 2.2 lists some of the popular univariate ranking schemes. While there exists many ranking schemes for 2-class discrimination problems, there are very limited number of ranking schemes for multi-class discrimination problems. Augmented Variance Ratio (AVR) [80, 81, 80, 88, 117, 120, 83, 119] is a less explored ranking scheme for multi-class discrimination. We show in Chapter 3,4 and 5 that AVR is a very discriminative ranking scheme. We also propose some new ranking schemes in this section and compare them with AVR and other popular ranking schemes (p-Value, Mutual Information, Corelation etc.). Any new ranking scheme can be easily be incorporated in the pipeline due to its modularity.
### Table 2.2: Popular Independent feature ranking schemes

<table>
<thead>
<tr>
<th>Classification</th>
<th>Regression</th>
</tr>
</thead>
<tbody>
<tr>
<td>Binary Class</td>
<td>2 or more discrete classes</td>
</tr>
<tr>
<td>p-Value [114]</td>
<td>VR [64, 15, 26]</td>
</tr>
<tr>
<td>Wilcoxon test [67]</td>
<td>WVR</td>
</tr>
<tr>
<td>Bhattacharyya distance [137]</td>
<td>WAVR</td>
</tr>
<tr>
<td>Receiver Operating Characteristics (ROC)</td>
<td>TAVR</td>
</tr>
</tbody>
</table>

#### 2.2.1 Variance Ratio (VR) [64, 15, 26]

For a feature 'f', the variance ratio is defined as the ration of total variance over the sum of variance of each class. Subscript ‘t’ stands for total while ‘C’ is the number of classes. Note that similarity of VR to fisher criterion.

\[
VR(f) = \frac{vr_t(f)}{\sum_{i=1}^{C} vr_i(f)} \tag{2.5}
\]

#### 2.2.2 Augmented Variance Ratio (AVR)[80, 88, 117, 120]

Augmented Variance Ratio (AVR) was first defined by Liu and Palmer [80]. For a feature f, AVR is the ratio of the variance of the feature between classes to the sum of the variance of the feature within classes, with an added penalty for features which may have small intra-class variance but which have close inter-class mean values. Subscript t stands for total while C is the number of
classes.

\[ AVR(f) = \frac{vr_i(f)}{\frac{1}{C} \sum_{i=1}^{C} \frac{vr_i(f)}{\{\min \neq j \mid \text{mean}_i(f) - \text{mean}_j(f)\}}} \]  (2.6)

### 2.2.3 Weighted Variance Ratio (WVR)

This ranking scheme similar to VR except that the variance of individual classes is weighted by number of samples, \( n_i \), in that class.

\[ WVR(f) = \frac{vr_i(f)}{\sum_{i=1}^{C} n_i vr_i(f)} \]  (2.7)

### 2.2.4 Weighted Augmented Variance Ratio (WAVR)

WAVR is similar to AVR except that instead of using the total variance in the denominator, it used the sum of central means weighted by number of subjects in each class. Thus it favors features whose mean is farthest from the central mean, with an added penalty for large intra-class variance.

\[ WAVR(f) = \frac{\sum_{i=1}^{C} n_i (\text{mean}_i(f) - \text{mean}_T)^2}{\frac{1}{C} \sum_{i=1}^{C} \frac{vr_i(f)}{\{\min \neq j \mid \text{mean}_i(f) - \text{mean}_j(f)\}}} \]  (2.8)

### 2.2.5 Total Augmented Variance Ratio (TAVR)

This is similar to AVR with a slight difference that, instead of weighting by distance between the closest 2 means, we reward features which are farthest
from the central mean. Unlike WAVR though, this scheme rewards features with high total variance.

\[
TAVR(f) = \frac{vr_i(f)}{\frac{1}{C} \sum_{i=1}^{C} \frac{vr_i(f)}{|\text{mean}_i(f) - \text{mean}_T|}} \tag{2.9}
\]

### 2.2.6 Total Weighted Augmented Variance Ratio (TWAVR)

TWAVR combines TAVR and WAVR. is similar to TAVR except that instead of using the total variance in the denominator, it used the sum of central means weighted by number of subjects in each class.

\[
nAVR - T(f) = \frac{\sum_{i=1}^{C} n_i (\text{mean}_i(f) - \text{mean}_T)^2}{\frac{1}{C} \sum_{i=1}^{C} \frac{vr_i(f)}{|\text{mean}_i(f) - \text{mean}_T|}} \tag{2.10}
\]

### 2.2.7 Weighted Ranking

Various ranking criterion described above can be combined as a weighed linear combination, to take advantage of different properties for different ranking schemes. Figure 2.3 compares the classification rate for some of the ranking schemes. Table 2.3 shows the discriminative regions selected by various ranking schemes for Chinese gender classification based on height feature, as described in section 4.4 of Chapter 4.
Figure 2.3: The classification rates averaged over 100 iterations when various feature ranking schemes were used. X-component of the deformation field, as described in section 5.4 of Chapter 5 was used for this experiment. After feature ranking, mRMR was used to select top 100 independent features, followed by SFS. LDA was used as the final classifier. It can be seen that AVR performed best. Correlation and Random ranking of features also performed relatively well, however their variance was found to be higher.
Table 2.3: Table showing the discriminative regions selected by various ranking schemes for Chinese gender classification (Section 4.2) based on height feature, as described in section 4.4. In the figure, the discriminative power of features varies from 'Red', which is highly discriminative, to 'Blue', which is least discriminative. A consistency between various ranking schemes can clearly be seen.
2.3 Redundancy Removal

After feature ranking we select top n\% of the features, where n can vary depending on the experiment. However many of these feature come from the similar region of the data (e.g. face/brain etc) and are highly correlated. This is especially true if the features have been extracted on a high resolution grid. Thus to remove redundant features while keeping only the relevant feature we further rank the feature subset by minimal-redundancy-maximal-relevance (mRMR)\cite{99}. The relevance is a measure of mutual information of an individual feature with the class label while the redundancy measure is based on mutual information of feature combinations. If the number of features is large (e.g. greater than 10,000) then finding of the maximal-relevance criterion becomes computationally intensive. This is the reason why we first select the top discriminative features based on AVR ranking followed by mRMR selection on this reduced subset rather than running mRMR on the whole feature set.

Note that other standard dimension reduction techniques like PCA can also be applied in place of/after mRMR ranking as the feature dimensions have already been reduced by feature ranking. However, as noted earlier in section 2.2, this would lead to a loss of individual feature correspondences.
2.4 Features Subset Selection

After features have been ranked by any of the criteria described in the previous section, we select the top 'n%' of the features. Optimal feature selection for supervised learning problems requires an exhaustive search of all possible subsets of features of the chosen cardinality. If large numbers of features are available (e.g. more than 1000), this is impractical. For practical supervised learning algorithms, the search is for a satisfactory set of features instead of an optimal set. In this study we use a wrapper-based feature selection algorithm [59]. Wrappers use a search algorithm to search through the space of possible features and evaluate each subset by running a model on the subset. While searching for a discriminative feature subset, we evaluate each subset by running a model (e.g. an LDA classifier) on the subset. In this process we optimize the classification rates for any one or more combination of classes. For example, in binary classification if classification rate is low due to low sensitivity, we can choose to optimize the the sensitivity and not the overall classification rate or an average of sensitivity and overall classification rate. The following different search strategies can be used for feature subset selection.
2.4.1 Sequential Forward Selection (SFS) [56]

Sequential forward selection starts with an empty feature subset. At each iteration exactly one feature is added to the feature subset. To determine which feature to add, the algorithm tentatively adds to the candidate feature subset one feature that is not already selected and tests the accuracy of a classifier built on the tentative feature subset. The feature that results in the highest accuracy is definitely added to the feature subset. The process stops after an iteration where no feature additions result in an improvement in accuracy.

2.4.2 Sequential Backward Selection (SBS) [56]

Sequential backwards selection is similar to forward selection, but it starts with all the features and attempts to eliminate iteratively a single feature that results in increase of accuracy gain. Although combination of features are taken into account with this technique, a high number of computations are necessary since it starts with the set of all features.

2.4.3 Take L - Remove R (LR) [33]

This is a generalization to SFS and SBS. Both in SFS and SBS only one feature is added or deleted at a stage. In LR search strategy, L features are added while R features are deleted at a stage. Thus SFS corresponds to LR with L=1 and
R=0 while SBS corresponds to LR with L=0 and R=1. In general LR versions are usually faster, since they examine fewer candidate feature subsets. Also L and R have to be specified by the user.

2.4.4 Sequential Floating Forward Selection (SFFS) [101]

The problem with SFS and SBS is its single-direction search. It tends to “go straight to the closest local optimum” of the search space. Thus it does not address the feature nesting problem - once a feature is added/removed, it remains so forever, even if the context changes and the feature could become important/useless inside some newly assembled subset. SFFS adds the functionality of backtracking and removing features after they have been added. However, the computational complexity is also increased.

2.5 Classification and Regression

In this stage we do the actual classification/Regression on test set using the features learned during feature selection stage. We use different classifiers such as Linear Discriminant Analysis (LDA) [26], Quadratic Discriminant Analysis (QDA) [113], Support Vector Machines (SVM) [125], K-Nearest Neighbors (KNN) [26] and Naive Bayes [26] for evaluating the performance on the test set.
Figure 2.4: The classification rates averaged over 100 iterations when various feature selection schemes were used. X-component of the deformation field, as described in section 5.4 of Chapter 5 was used for this experiment. After AVR feature ranking, mRMR was used to select top 100 independent features, followed by SFS. LDA was used as the final classifier.

In summary, the classification pipeline consists of the following steps (Figure 2.1):

a) Divide the dataset randomly into test and training sets.

b) Normalize each feature in Training set based on any ranking scheme described in Section 2.1. Scale the Test set using the same scale that was used to scale the Training set.

c) Rank the features in training set based on one of the criterion discussed in the Section 2.2 and select the top ‘n%’ of the features.

d) Do mRMR on this reduced feature subset to remove redundancy.
Figure 2.5: The classification rates averaged over 100 iterations when various optimization criterion were used. X-component of the deformation field, as described in section 5.4 of Chapter 5 was used for this experiment. After feature ranking, mRMR was used to select top 100 independent features, followed by SFS. LDA was used as the final classifier.

e) Do feature subset selection on this reduced feature subset. The model evaluation at each iteration in the selection process can be done on the training set or the validation set.

f) Either directly select the top features returned after step (e) or repeat step (a) - (e), fixed number of times and select the “best subset” of features based on the frequency with which each feature is selected in the given number of repetitions or on its average performance on the validation set. To reduce variability the process is repeated multiple times with different partitions. This method is applied especially when we have low training samples and the model
Figure 2.6: The classification rates averaged over 100 iterations when different classifiers were used as wrappers during feature selection. X-component of the deformation field, as described in section 5.4 of Chapter 5 was used for this experiment. After AVR feature ranking, mRMR was used to select top 100 independent features, followed by SFS. LDA was used as the final classifier.

is overfitting [89].

g) Repeat steps (a) - (f) N times and compute the average statistics.
Figure 2.7: The classification rates averaged over 100 iterations when Classifiers were used. X-component of the deformation field, as described in section 5.4 of Chapter 5 was used for this experiment. After AVR feature ranking, mRMR was used to select top 100 independent features, followed by SFS.

<table>
<thead>
<tr>
<th></th>
<th>MAH</th>
<th>KNN (K=1)</th>
<th>LDA</th>
<th>QDA</th>
<th>SVM</th>
</tr>
</thead>
<tbody>
<tr>
<td>MAH</td>
<td>68.3(±4.7)</td>
<td>61.1(±5.5)</td>
<td><strong>68.4(±5.0)</strong></td>
<td>68.2(±4.7)</td>
<td>66.2(±5.2)</td>
</tr>
<tr>
<td>KNN (k=1)</td>
<td>61.3(±5.6)</td>
<td>57.8(±5.0)</td>
<td>61.8(±5.3)</td>
<td>61.1(±4.3)</td>
<td><strong>61.6(±5.1)</strong></td>
</tr>
<tr>
<td>LDA</td>
<td>70.5(±4.0)</td>
<td>64.3(±5.8)</td>
<td><strong>70.9(±4.4)</strong></td>
<td>70.6(±4.5)</td>
<td>68.7(±3.0)</td>
</tr>
<tr>
<td>QDA</td>
<td>68.7(±4.8)</td>
<td>63.1(±4.6)</td>
<td>68.8(±4.3)</td>
<td><strong>69.0(±4.3)</strong></td>
<td>66.3(±3.6)</td>
</tr>
<tr>
<td>SVM</td>
<td><strong>67.0(±3.4)</strong></td>
<td>61.1(±5.6)</td>
<td>66.4(±3.9)</td>
<td>66.3(±3.6)</td>
<td>65.6(±4.3)</td>
</tr>
</tbody>
</table>

Table 2.4: Comparison of different classifiers as wrappers for feature selection. Each row indicates the classifier that was used for feature selection, while each column indicates the final classification rate using different classifiers. The best classification in each row has been highlighted. X-component of the deformation field, as described in section 5.4 of Chapter 5 was used for this experiment. After AVR feature ranking, mRMR was used to select top 100 independent features, followed by SFS. The best result was achieved when LDA was used as a wrapper as well as the final classifier. NOTE: MAH = Mahalanobis, KNN= K-nearest neighbor, LDA = Linear Discriminant Analysis, QDA = Quadratic Discriminant Analysis, SVM= Support Vector Machine (linear kernel)
A major field of research in plant study is associated with identifying genes which can make plant requiring less resources, become more productive and resistant to environmental factors [6, 61]. Advances in genetic technology over the last decade has enabled scientists to map the genome sequence of plants and make custom modifications in the gene-sequence [44, 6, 61]. The goal of these studies is to link a genotype (a genetic sequence) to a phenotype (morphological and other physical characteristics of plant which can be observed). However, many times the genetic modifications in the plant lead to very subtle changes which are difficult to discriminate visually by a human observer. This makes it difficult to study the effect of the gene-modification. More often these changes are relatively easily identified in the leaves than in other parts of the plant.
Thus automatic feature extraction and computational analysis from leaves form an obvious first step in understanding of plants. On a macroscopic level, scientist have also tried to identify plant species from features extracted from leaves. This Chapter gives a summary of our work done in this field using a two datasets as described below.

3.1 Leaf Datasets

We use two datasets in this work. One is the publicly available benchmark dataset of 1125 Swedish leaves provided by E. Keogh [53] and the other is the ecotypes dataset provided by Prof Hong Ma from Eberly College of Science, The Pennsylvania State university.

3.1.1 Swedish leaf dataset

The Swedish leaf dataset provided by [53] has leaves from 15 plant species with 75 samples each. It was originally collected by Swedish Museum of Natural History to build a computer based automatic species classifier in 1998. Figure 3.1 shows a sample set of leaves from each of the 15 classes.
Figure 3.1: Sample leaves from 15 species of the Swedish leaf dataset. Note the similarity between some classes.

### 3.1.2 Ecotypes Dataset

The ecotypes dataset is provided by Prof Hong Ma from Eberly College of Science, The Pennsylvania State University. The dataset contains two ecotypes of the plant Arabidopsis thaliana. Although both ecotypes come from the same species, they come from different parts of the world. Hence they have different phenotype (morphology) and genetic background. Identification of features which discriminate these ecotypes, would help in deeper understanding of the
genetic basis for leaf shape control, i.e. what genetic factors control the shape of the leaf.

The dataset contains 10 samples for each dataset. For each sample, the plant leaves are imaged after 24'th day of growth. The leaves are arranged in order of their age as shown in figure 3.2. The leaf age decreases linearly from left to right and top to bottom. The oldest leaf appear on top of the plant while the youngest appear near the bottom. Since the youngest leaf is generally not fully grown at the 24th day of growth and the oldest leaf often loses some typical characteristics, so we choose to processes only 6 set of leaves ( shown inside blue bounding box in figure 3.2 ).

3.2 Related work

The first work on the Swedish dataset was reported by Oskar Söderkvist [111]. By extracting features such as area, circularity, Flusser-moments, Hu-moments, eccentricity, curvature scale space, incremental circle transform and using a neural network classifier, he reported a best classification rate of 82.4%. It was also reported that tree class 1 was often misclassified as tree class 3, 6, 9, 12 and 15, and tree class 14 as 4, 6 and 13. Ling et al [68] proposed using the inner-distance as an extension of Shape Context (SC) [10] which they refer to as IDSC. They reported a classification rate of 89.60% using Fourier de-
Figure 3.2: A Sample from the ecotypes dataset. The blue rectangle encloses the leaves of interest that we study.

scriptors, 88.12% using SC + Dynamic programming and 94.13% using IDSC+ Dynamic Programming respectively. 1-nearest neighbour was used as a classifier. They further extended their work in [69] by using Shortest Path Texture Context (SPTC) + Dynamic Programming and achieved a best classification rate of 95.33%. Similar results were obtained by using Multi-Dimension Scaling (MDS) + Shape Context + Dynamic Programming. Felzenszwalb and Schwartz [31] proposed a new hierarchial representation for curves called shape tree and achieved a best classification of 96.28% on the swedish leaf dataset. Xi et al[136] showed that the swedish leaf dataset can be converted into a “pseudo time se-
ries” by measuring the distance between the contour points and the centroid. They also proposed a new algorithm based on 1 nearest neighbour classifier with Dynamic Time Warping (DWT) called Adaptive WARPing winDow (AWARD) for time-series classification. They achieved a best classification of around 85% by applying the AWARD algorithm to the swedish leaf dataset. Wei et al.[131] also treated the swedish leaf dataset as a time series by converting shapes to one dimensional representations using centroid distance. By using a 1-nearest neighbour classifier with rotation invariant distance, they report a best classification rate of 86.67%. Sahbi [106] reported a best precision rate of 96.7% by using Kernel PCA. Zhang et al.[142] proposed a new clustering algorithm called StrAP (streaming affinity propagation) which represented the dataset using a message passing method. By using StrAP, they reported a classification rate of 94.3%. Wu and Rehg [135] proposed a new descriptor called PACT (Principal component Analysis of Census Transform histograms) and report a best classification 90.77% using only shape and that of 97.92% using both shape and intensity. They used SVM classifier with a RBF kernel. By modelling shape curves as elastic strings [129] propose an extension of homogeneous elastic models (HEM) to planar curves and report a best classification rate of 96.67% on the swedish leaf dataset. Bai et al.[8] considered existing shapes as a group and studied their similarity measures to query shape in a graph structure. They report a best classification rate of 93.8%. Wachman [126, 127] treated the Swedish
leaf as a time series by plotting the radius of a line anchored in the center of
the object as it rotates around the image. They ran a 10-fold cross-validation
using different classifiers and report a best classification rate of $82\pm3\%$ using 1
nearest neighbour kernal classifier.

3.2.1 Choice of the Swedish leaf dataset

As described in the previous section, researchers have applied a wide variety of
methods to classify plant species in the Swedish leaf dataset and achieved above
90% classification accuracy. However we still choose this dataset because of the
following two reasons:

1. The performance of the classification pipeline that we propose in this work
   has to be compared to some existing benchmark dataset. Thus Swedish
   leaf dataset becomes a natural choice as many different algorithms have
   been applied on this dataset. Thus competitive classification results on this
dataset would demonstrate the robustness of our classification pipeline.

2. The second and more important reason is to get an insight into the leaf
   properties. While automated classification of species may have practical
   applications, we should go further and ask what makes the plant species
different. Why are some leaves more curved than others ? What controls
the color of leaves ? How much role does environment play in the leaf
<table>
<thead>
<tr>
<th>Author</th>
<th>Year</th>
<th>Method</th>
<th>Feature</th>
<th>Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>Oskar Soderkvist [111]</td>
<td>2001</td>
<td>Neural Network</td>
<td>area, circularity, Flusser-moments, Hu-moments, eccentricity, curvature scale space, incremental circle transform</td>
<td>82.4%</td>
</tr>
<tr>
<td>Ling et al [68]</td>
<td>2005</td>
<td>1-nearest neighbour</td>
<td>Fourier descriptors</td>
<td>89.60%</td>
</tr>
<tr>
<td>Ling et al [68]</td>
<td>2005</td>
<td>1-nearest neighbour</td>
<td>Shape Context (SC) + Dynamic programming</td>
<td>88.12%</td>
</tr>
<tr>
<td>Ling et al [68]</td>
<td>2005</td>
<td>1-nearest neighbour</td>
<td>Inner-Distance Shape Context (IDSC) + Dynamic programming</td>
<td>94.13%</td>
</tr>
<tr>
<td>Xi et al [136]</td>
<td>2006</td>
<td>Adaptive WRaping winDow (AWARD) (based on 1 neared neighbor)</td>
<td>Boundary signature converted to “pseudo time series”</td>
<td>85%</td>
</tr>
<tr>
<td>Wei et al [131]</td>
<td>2006</td>
<td>1-nearest neighbour classifier</td>
<td>pseudo-time series with rotation invariant distance from centroid</td>
<td>86.67%</td>
</tr>
<tr>
<td>Ling et al [69]</td>
<td>2007</td>
<td>1-nearest neighbour</td>
<td>Shortest Path Texture Context (SPTC) + Dynamic Programming</td>
<td>95.33%</td>
</tr>
<tr>
<td>Ling et al [69]</td>
<td>2007</td>
<td>1-nearest neighbour</td>
<td>Multi-Dimension Scaling (MDS) + Shape Context + Dynamic Programming</td>
<td>95.33%</td>
</tr>
<tr>
<td>Sahbi [106]</td>
<td>2007</td>
<td>Kernel PCA</td>
<td>Multi-Dimension Scaling (MDS) + Shape Context + Dynamic Programming</td>
<td>95.33%</td>
</tr>
<tr>
<td>Felzenswalb and Schwartz [31]</td>
<td>2007</td>
<td>Shape Matching</td>
<td>Shape Tree</td>
<td>96.28%</td>
</tr>
<tr>
<td>Wachman [126, 127]</td>
<td>2009</td>
<td>1 nearest neighbour kernal classifier</td>
<td>pseudo-time-series</td>
<td>82%</td>
</tr>
</tbody>
</table>

Table 3.1: Selected work on plant specie Classification
characteristics? To answer question such as these we need to identify features which have some physical meaning. We should also study both inter and intra-distribution of such features. The long term goal should be to identify genes which control these features. In this work, we attempt to identify discriminating features between species and study its variations, besides classification as done by other researchers.

3.3 Preprocessing

The leaf has to be segmented from the background before features can be extracted from it. This task is relatively easy as the images in both the Swedish and ecotypes datasets have uniform backgrounds. We treat this task as a pixel classification problem. First we randomly select patches from both foreground and background and then find the mean of the RGB image channels for both. A pixel is classified as foreground or background depending on its euclidean distance from the mean RGB values. Figure shows the template which is used to calculate the mean RGB pixel.

For the ecotypes datasets, the six leaves needed for processing are identified by the position of the centroid of the blobs after segmentation. The whole process is illustrated in figure 3.4.
Figure 3.3: Segmentation templates for leaf segmentation. (a) and (c) show the background template for ecotypes and Swedish datasets respectively. (b) and (d) show the corresponding foreground template.

Figure 3.4: Preprocessing stages for ecotypes dataset. (a) Original Image (b) Classified Foreground (c) Segmentation Mask (d) Identified leaves
3.4 Feature Extraction

A leaf can be characterized by its size, shape (linear/ovate/cordate etc), margin (shape of edge, eg. entire/tothed/lobed), venation (longitudinal/palmate/arcuate etc) and its texture. For more information about leaf morphology refer to [132]. Some of the leaf properties such as area can be easily calculated while others such as venation (pattern of veins in the blade of a leaf) are difficult to accurately characterize. The goal of the feature extraction process is to capture a set of features which compactly and comprehensively characterize a leaf and also robustly calculated. The subsections below list the features that we use in this work. The features have been divided into two groups, the ones extracted from shape and the ones extracted from texture. For many of these features we also extract a combination of statistical moments and features derived from moments as described in the next subsection.

3.4.1 Features from Shape

Shape is an important leaf characteristics. This has also been the most popular among plant physiologists as some of its properties are relatively easy to measure. The following subsections describe the shape features that we use in this work.

**Area** After the preprocessing step, we get a binary segmentation mask indi-
Figure 3.5: Some basic leaf properties. The leaf contour is shown in red. The blue rectangle is the bounding box. The blue ‘plus’ in the middle of the leaf is the centroid. The figure on the right shows the boundary signature [93]. Zero degree corresponds to horizontal right, with theta increasing in counter clockwise direction.

cating the foreground and background. Area is the the sum of number of foreground pixels. This is indicated by the white region in figure 3.5.

**Perimeter** The perimeter is calculated as the distance between each adjoining pair of pixels around the border of the region.

**Compactness** This is defined as \( \text{perimeter}^2 / \text{area} \) and is insensitive to uniform scale changes. Compactness is minimal for a disk shaped region.

**Bounding box area** This is the area of the smallest rectangle enclosing the leaf contour.
**Centroid**  This is the co-ordinate of the center of mass of the leaf relative to the left-top corner of the bounding box.

**Solidity**  This is the ration of area to number of pixels in the convex hull (convex area).

**Eccentricity**  This is the eccentricity of the ellipse having same normalized second-central moments as the region.

**Major Axis Length**  This is the length of major axis of the ellipse having same normalized second-central moments as the region. This can be interpreted as length of the leaf.

**Equivalent Diameter**  This is the diameter of the circle having same area as the leaf.

**Minor Axis Length**  This is the length of minor axis of the ellipse having same normalized second-central moments as the region. This can be interpreted as width of the leaf.

**Moments of Boundary Signature**  Boundary signature is the distance of the boundary from the centroid as function of angle. Figure 3.5 shows the boundary signature of a leaf from Swedish dataset. It can be seen that it is a simple yet powerful descriptor accurately capturing the leaf contour. However, the boundary signature is sensitive to noise and is not directly comparable across leaves. Thus to characterize the boundary signature
Table 3.2: A set of 20 statistical features used for plant feature extraction. See Appendix I for more details

robustly we extract a list of 20 statistical features from it. These are listed in Table 3.2. For more information about moments, refer appendix 1 or [20, 36].

**Moments of Boundary curvature** Curvature of a 2D curve is the inverse of the radius of a circle that is tangent to that curve at a particular location [57]. This is defined mathematically by equation 3.1

\[
\kappa = \frac{x' y'' - y' x''}{(x'^2 + y'^2)^{3/2}}
\]  

(3.1)

Figure 3.6 shows the curvature of leaf plotted along the leaf contour for a leaf from the Swedish dataset. After we find the curvature, we extract 20 statistical moments, as listed in Table 3.2. These are used as a compact representation of the curvature.

**Fourier Descriptors** The boundary of the leaf is a collection of \([x, y]\) coordinates, where \(N\) is the number of points along the boundary. By interpreting the X-axis as the real axis and Y axis as imaginary axis, the contour of the leaf can be treated as

<table>
<thead>
<tr>
<th>Mean</th>
<th>Variance</th>
<th>Skewness</th>
<th>Kurtosis</th>
<th>Moment 6</th>
<th>Median</th>
<th>Min</th>
<th>Max</th>
<th>Smoothness</th>
</tr>
</thead>
<tbody>
<tr>
<td>Moment 6</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Contrast</td>
<td>Uniformity</td>
<td>Entropy</td>
<td>7 Hu-Moment Invariants [47]</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>


Figure 3.6: Different views of curvature of a leaf plotted along the boundary.

\[ s(k) = x(k) + iy(k) \quad \text{for} \quad k = 0, 1, \ldots, N - 1 \quad (3.2) \]

We can now find the Discrete Fourier Transform (DFT) of \( s(k) \) as given by equation 3.3.

\[ a(u) = \frac{1}{N} \sum_{n=0}^{N-1} s(k)e^{-\frac{i2\pi uk}{N}} \quad (3.3) \]

We then find moments of the phase and amplitude response of the DFT of the fourier descriptor. Figure 3.7 shows the amplitude and phase response of the the leaf boundary shown is figure 3.5.
3.4.2 Features from Texture

Leaf color and texture is a very important feature. Traditionally they have not been used by plant physiologists as they were difficult to characterize manually. However this can easily be done from digital leaf images. We list below the texture features that we use in this study.

Histogram of color channels Histogram of a digital image represents its tonal distributions. A digital color image has three channels, one each for red, green and blue. We evaluate 64 bin histogram for each of these channels. Further we transform the image from RGB color space to CIELAB color
space. CIELAB color space is designed to approximate human vision unlike RGB color space. We then calculate 64 bin histograms in each of the L, A, B channels. Figure 3.8 shows the process for RGB channels.

**Moments of color channels and histograms** We also find 20 statistical moments of the histograms as well as texture. The utility of finding such moments is illustrated in figure 3.9. The first columns shows 5 different leaf textures. The second column shows the corresponding RGB histogram. Finally the third column lists a subset of moments. It can be seen that to discriminate Image 3 form Image 1, skewness can be an important feature. Similarly entropy can be an important feature to discriminate between Image 1 and Image 2.

**Moments of Spectral Measures** Fourier spectrum is a handy tool to measure regularity in a texture. In the context of leaves, the arrangement of
veins or a regular stripe pattern on leaves can be easily distinguished by
the fourier analysis. First we calculate the fourier spectrum of the texture
and convert it to polar co-ordinates $S(r, \theta)$. Summation of the response for
all $\theta$ gives a 1D function $g S(r)$ while summation for all $r$ gives $S(\theta)$. To
compactly characterize the spectrum we find 20 statistical measures each
for $S(r, \theta)$, $S(r)$ and $S(\theta)$. Figure 3.10 shows some of these features for 3
leaves from the Swedish leaf dataset.
Figure 3.10: Spectral features for 3 leaves from Swedish dataset

**GLCM Texture measures** The Grey Level Co-occurrence Matrix (GLCM), also called the Grey Tone Spatial Dependency Matrix, is a tabulation of how often different combinations of pixel brightness values (grey levels) occur in an image. It captures the spatial distribution of intensity variations in each image. Each element \((i,j)\) in GLCM specifies the number of times that the pixel with value \(i\) occurred adjacent to a pixel with value \(j\). We measure the adjacency in 4 directions, 0, 45, 90 and 135 degrees. Before calculations, the values in the image is scaled to 8 gray-levels. After cal-
The calculation of the GLCM matrix, it is normalized so that its values sum to 1. Thus each element \((i,j)\) in the normalized GLCM is the joint probability occurrence of pixel pairs with a defined spatial relationship having gray level values \(i\) and \(j\) in the image.

### 3.5 Results and Discussions

We performed discrimination experiments on two datasets, the Swedish leaf dataset and ecotypes.

#### 3.5.1 Swedish Leaf Dataset

Unlike ecotypes dataset, species classification from Swedish leaves is a multi-class problem. Besides the primary goal of building an automated system for plant species classification from leaf images, we are also interested in identifying features which are responsible species differentiation. This is of special interest to genetic-biologists who link morphological traits to genes as it would give insight into species evolution.

The set up of the experiment is similar to that on the ecotypes dataset. In this dataset we have 75 samples from each class and 15 classes in total. We use 50 samples from each class for training and 25 for testing. This distribution of samples is similar to many of the studies on this dataset as described in Section
Table 3.3: Average classification for Swedish leaf rates over 100 random splits.

<table>
<thead>
<tr>
<th></th>
<th>Texture Only</th>
<th>Shape only</th>
<th>Texture + Shape</th>
</tr>
</thead>
<tbody>
<tr>
<td>Classification rates</td>
<td>97.72(±5.5)</td>
<td>93.33(±4.48)</td>
<td>98.93(±2.3)</td>
</tr>
</tbody>
</table>

3.2. As the feature dimension was less than 10000, so we excluded the AVR ranking. Thus for this experiment we used a mRMR + SFS feature selection is then applied to 100 random splits of the dataset. LDA was used as the final classifier. However to find important features AVR was done independent of the classification process. The was done because the mRMR algorithm would remove redundant features even though it may be discriminative individually. This makes sense for classification purposes, but not for individual feature importance (although we could use only the “MR” i.e. maximum relevance part of the mRMR algorithm also to gage individual feature importance).

Table 3.5 lists the classification rates when only shape, only texture and both shape and texture are used. All the results are reported after averaging over 100 random splits. The classification rate of 98.93 ± 2.3 % that we report is the highest reported so far on this dataset and slightly better than the previous best 97.92 % reported by Wu and Rehg [135]. As can be seen from table 3.5, texture features play a more significant role than shape feature.
3.5.2 Ecotypes Dataset

The aim of experimentation on the ecotypes datasets is to better understand how environment effects plants of same species. This dataset has 2 ecotypes of the same species but collected from different parts of the world. The ecotypes are named Graz-1 and Dizon-1. Each sample contains 2D color image of 6 leaves as described in chapter 2. The feature dimension when all the features described in chapter 2 are calculated is 3885. Each of the six leaf has a special temporal meaning (they differ in age). Hence they have to be considered as a group. Thus the effective feature dimension becomes $6 \times 3885 = 23310$.

For classification purpose, we first divide the dataset into test and training set such that 9 samples from each class is used for training and 1 from each class is used for testing. First we scale the features to $[0-1]$. We then rank the features based on AVR followed by mRMR redundancy removal. SFS is then applied on a reduced feature subset of 100 features. The mean of sensitivity, specificity and classification rate is optimized during the feature selection. LDA is used as a wrapper. The process is repeated 100 times each with random splitting into test and training set. Figures 3.12 and 3.13 show the feature distribution in test and training set when only texture and shape features were used for classification. Table 3.5 summarizes the classification results.

It was found that texture features were selected more often than shape features.
<table>
<thead>
<tr>
<th>Index</th>
<th>Feature</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>5th leaf Entropy of histogram of green channel</td>
</tr>
<tr>
<td>2</td>
<td>4rd Leaf moment 3 texture of green channel</td>
</tr>
<tr>
<td>3</td>
<td>2’nd leaf Leaf Area</td>
</tr>
<tr>
<td>4</td>
<td>3’rd leaf 4th moment 3 of Green texture (G component, RGB space)</td>
</tr>
<tr>
<td>5</td>
<td>3’rd leaf 5th moment invariant of Green texture (G component, RGB space)</td>
</tr>
<tr>
<td>6</td>
<td>5’th leaf 5th moment invariant of Green texture (L component, LAB space)</td>
</tr>
<tr>
<td>7</td>
<td>4’th leaf 6th moment invariant of Green texture (L component, LAB space)</td>
</tr>
<tr>
<td>8</td>
<td>4’th leaf 6th moment invariant of Green texture (G component, RGB space)</td>
</tr>
<tr>
<td>9</td>
<td>5th leaf Median of Spectral response</td>
</tr>
<tr>
<td>10</td>
<td>1st leaf 3rd moment invariant of Green texture (L component, LAB space)</td>
</tr>
</tbody>
</table>

Table 3.4: Top 10 AVR ranked features

<table>
<thead>
<tr>
<th></th>
<th>Texture Only</th>
<th>Shape only</th>
<th>Texture + Shape</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dizon-1 (Sensitivity)</td>
<td>100 (± 0)</td>
<td>80 (± 32.1)</td>
<td>90 (± 15.6)</td>
</tr>
<tr>
<td>Graz-1 (Specificity)</td>
<td>80 (± 32.1)</td>
<td>100 (± 0)</td>
<td>100 (± 0)</td>
</tr>
<tr>
<td>Overall</td>
<td>90 (± 15.0)</td>
<td>90 (± 16.0)</td>
<td>95 (± 7.8)</td>
</tr>
</tbody>
</table>

Table 3.5: Summary of classification experiments on Ecotypes dataset

Moments of histogram of texture were mostly selected. For shape based features perimeter, extent and area were found to be most important for ecotypes discrimination. It can be observed from the classification table that texture features are more important in identifying Dijon-1, while shape features identify Graz-1 easily.
Figure 3.11: Confusion matrix for Swedish leaf classification. The first column is training while the second column is test. All results are averaged over 100 iterations.
Figure 3.12: Feature distribution in top 2 selected dimension during a random split. Only texture features were used.

Figure 3.13: Feature distribution in top 2 selected feature dimension during a random split. Only shape features were used.
Figure 3.14: Feature distribution in top 3 AVR ranked feature
Chapter 4

3D Face Discrimination

Human perception and discrimination of gender, expression, race, age etc. has been investigated from both psychological and computational perspectives [19, 2, 16, 1, 40, 7]. Despite the advances in face recognition technology over the last few decades, very little is known about the neural mechanisms of the human brain that discriminate gender, expression etc. Questions such as which part of the face is important for gender or face recognition, remain unanswered. Problems such as gender, face or expression recognition, while almost trivial for a human observer, pose considerable challenge to a computer. Traditionally most of the work on human facial analysis has been in 2D, over the last decade however the focus is rapidly shifting from 2D to 3D. This is primarily because of the fact that problems such as pose changes, lighting variation, subtle facial
movements etc, which form a major stumbling block in 2D processing, can be easily handled in 3D. The rapid advancement of 3D imaging technology and reducing costs have also played a major role in this shift. 3D approach can also help in detection and analysis of subtle facial movements, which could lead to a range of new applications such as lie detection, something that is not possible by using 2D data only. Study of faces based on 3D data would also be a step closer towards understanding of human perception and cognition.

In this chapter we first review the work done on 3D facial analysis.

### 4.1 Related work

One of the earliest papers to use truly 3D data for facial analysis was presented by Gordon [37, 38]. He used both 3D range data and facial surface curvature for face recognition. However many of the works that followed used only the range data. Most of the early work in the field of 3D facial analysis were limited to face recognition. It is only in the recent years that researchers have started directing their effort to other aspects of face analysis such as extracting information related to gender, expression, age, race etc from 3D range data.
4.1.1 3D Gender recognition from face

It was demonstrated by Toole et al [96] that gender classification is better in 3D range data than with image intensity information. The usefulness of 3D data for gender recognition was further demonstrated by Liu and Palmer [80]. By making use of height and facial orientation based 3D facial (a)symmetry features, they reported a best classification rate of 96% on a 3D dataset containing 69 males and 36 females. Lu et al [84] used 3D range data and texture information for gender classification of 376 subjects to achieve a best classification rate of 91%. However, the classification rate dropped to 86% when only range data was used. Also their results were not symmetric in terms of gender classification with mean error rate of 24.5% for females and 9% for males. There has not been any recent work in 3D gender recognition using 3D exclusively.

4.1.2 3D Expression recognition from face

3D expression recognition has become an active area of research in recent years. Yabui et al [138] were one of the first to use 3D range data for expression classification. They reported a best rate of 63.44%. Wang et al [128] proposed an algorithm where expressions are classified according to the distribution of several surface geometry descriptors in seven expressive facial regions. These regions are defined according to the neuro-anatomy knowledge of configuration.
of facial muscles and their dynamics. The combination of seven histogram distributions generates a unique expression descriptor for a specific expression. They report a best classification rate of 83.6% on BU-3DFEDB database using Linear Discriminant Analysis classifier. Soyel and Demirel [112] proposed a Neural network classifier to recognize the facial expressions from a distance vector obtained from 3D 83 manually selected facial feature locations to achieve a best classification rate of 91.3% on the BU-3DFEDB database. Mpiperis et al [91] reported an overall classification rate of 92.3% on the same database using a classifier based on a set of rules which are discovered following the principles of swarm intelligence, in particular of Ant Colony Optimization (ACO) and Particle Swarm Optimization (PSO). Mpiperis et al [92] also proposed bilinear models for jointly addressing 3D face and facial expression recognition and use it to achieve a overall classification rate of 90.5% on BU-3DFE database. The best classification rate so far has been reported by Tang and Huang [116]. They use a combination of average relative entropy based feature selection algorithm to select features composed of normalized Euclidean distances between 83 facial feature points in the 3D space. By using a regularized multi-class AdaBoost classification algorithm they report a 95.1% average classification rate.
4.1.3 3D face recognition under expression variations

Most of the early work in the field of 3D facial analysis assume a neutral face. In reality however this assumption is very restrictive. Some of the work which address the problem of 3D face identification under expression variation follow a multi-region technique. First the face is divided into various regions. To account for expression variations, features are chosen from areas which remain unaffected by facial expressions, such as eyes and nose. [98, 52] use an Annotated Face Model which is deformed elastically to fit each face thus allowing the annotation of its different anatomical areas like nose, eyes, mouth, etc. Chang et al[21] follow a multi-region technique in which multiple overlapping regions around the nose are matched using the Iterative Closest Point algorithm (ICP). These approaches suffer from the problem that parts of the face having discriminative information are rejected during matching. Li et al. [65] classify faces using expression-invariant descriptors that are based on surface curvature, geodesic distance and attributes of a 3D mesh fitted to the face. The descriptors are weighted appropriately to form a final face signature that is used to identify faces.

Table 4.1 provides a succinct summary of some recent work on 3D face discrimination (identity, gender, expression). Though classification results using different methods on various image data sets are reported, little is reported for
<table>
<thead>
<tr>
<th>Year</th>
<th>Dataset</th>
<th>Feature Description</th>
<th>Method</th>
<th>Rates</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gordon [37]</td>
<td>1991</td>
<td>3D range data for 24 subjects</td>
<td>Distance and curvature measures of some manually selected regions e.g. gaussian curvature at bridge of the nose etc.</td>
<td>Non-discriminative features based on fisher score are discarded. Classification by euclidean distance in the remaining feature space.</td>
</tr>
<tr>
<td>Passalis et al [98], Kakadiaris et al [52]</td>
<td>2005, 2007</td>
<td>FRGC v2 (884 subjects)</td>
<td>Deformable Annotated Face Model + wavelet coefficients</td>
<td>Variation of the CW-SSIM to measure distance metric used for retrieval</td>
</tr>
<tr>
<td>Cheng et al [21]</td>
<td>2006</td>
<td>UND (446 subjects/4000 scans)</td>
<td>Multiple nose region matching. Overlapping regions around the nose are matched using the Iterative Closest Point algorithm (ICP).</td>
<td>Sum, product and minimum rule to combine similarity scores which is used for retrieval</td>
</tr>
<tr>
<td>Li et al. [65]</td>
<td>2007</td>
<td>GavabDB (300 subjects) / FRGC 3D (180 subjects)</td>
<td>expression-invariant descriptors that are based on surface curvature, geodesic distance and attributes of a 3D mesh fitted to the face.</td>
<td>PCA + Nearest neighbor classification</td>
</tr>
<tr>
<td>Liu and Palmer [80]</td>
<td>2003</td>
<td>USF (111 subjects)</td>
<td>Height and orientation facial asymmetry</td>
<td>AVR + LDA (100 iterations)</td>
</tr>
<tr>
<td>Wang et al [128]</td>
<td>2006</td>
<td>BU3DFE (60 subjects)</td>
<td>Distribution of several surface geometry descriptors in seven expressive facial regions.</td>
<td>LDA (20 iterations)</td>
</tr>
<tr>
<td>Soyel and Demirel [112]</td>
<td>2007</td>
<td>BU3DFE (60 subjects)</td>
<td>Distance vector obtained from 3D 83 manually selected facial feature locations</td>
<td>Neural network classifier (10 iterations)</td>
</tr>
<tr>
<td>Tang and Huang [116]</td>
<td>2008</td>
<td>BU3DFE (60 subjects)</td>
<td>24 manually defined features based on distance between key fiducial points</td>
<td>Combination of average relative entropy based feature selection algorithm + regularized multi-class AdaBoost (10 iterations)</td>
</tr>
<tr>
<td>Mpiperis et al [91]</td>
<td>2008</td>
<td>BU3DFE (100 subjects)</td>
<td>Range data</td>
<td>PCA + LDA to get a 5D vector. Ant Colony Optimization (ACO) and Particle Swarm Optimization (PSO) (10 iterations)</td>
</tr>
<tr>
<td>Mpiperis et al [92]</td>
<td>2008</td>
<td>BU3DFE (100 subjects)</td>
<td>Elastically deformable model</td>
<td>Bilinear models for jointly addressing 3D face and facial expression recognition (10 iterations)</td>
</tr>
</tbody>
</table>

Table 4.1: Some of the recent work in 3D face discrimination
repeated tests on the same data set and using the same number of subjects for training/testing with a sufficient amount of statistics (report both the mean and the standard deviations of the classification rates from multiple tests).

4.1.4 Facial asymmetry

It has been reported that removing asymmetry from facial images leads to a small but significant drop in facial recognition rate [94]. Studies of attractiveness have also reported preferences for symmetrical faces [104, 30]. It has also been reported that male faces are more asymmetrical than female face. Considering the above information (a)symmetry profile of human face becomes the natural choice to be used as a feature for gender classification. The use of (a)symmetry information for gender classification was first proposed by Liu and Palmer [80]. Facial asymmetry has also been successfully applied to 2D expression classification [88, 118]. There is no corresponding work in 3D which uses facial asymmetry for expression classification.

4.1.5 Shortcomings of existing approaches

While there are a number of paper on 3D expression classification which report a classification rate in the range of 80 to 95% , most of them rely on a geometric relations between some key fiducial points. In particular [128, 112, 91, 92,
use 83 key fiducial points for feature extraction. These points have been manually labeled in the BU3DFE database [139]. Many of these points lie on parts of the face where there are no major landmarks, for example at end of the cheeks. Such manual labeling is subjective and is a source of error. It also involves a lot of effort to mark such large number of points for large datasets. In this work, we only use 3 fiducial points for face normalization. Moreover, once the faces have been normalized, the feature extraction process does not depend on the location of these points. Although in the current work we manually label these three points, there are algorithms which can automatically detect these three points without disambiguity, hence automating the whole process [23]. Another drawback of existing methods is that the features used are not generic in the sense that the same set of features may not be applied to other aspects of facial analysis, apart from which it was designed for. For example most existing methods for 3D expression classification rely of euclidian distance between some key fiducial points. These methods cannot be applied to gender recognition, where the location of these fiducial points is almost same across gender. Also methods relying only on distance between fiducial points would fail to detect subtle facial changes. Besides, only 3D range data is not enough to fully describe a 3D surface. Surface curvature and orientation should also be included when analyzing a 3D surfaces. Also most of the work on 3D expression and gender classification use features that are global. While, they may produce
resonable classification rates, but they cannot identify local regions of the face which are most important for a particular discrimination task.

4.2 3D Face Datasets

The next subsection described the 3D face datasets that we use in this study.

4.2.1 BJUT-3DR1 dataset (500 subjects)

For gender classification we use the BJUT-3D face database [14]. It has 250 Chinese male and 250 Chinese female subjects. All the 3D faces in the BJUT-3DR1 database have neutral facial expression and are without glasses or other accessories. Each image file is stored in the database in the form of a triangle mesh, which consists of X, Y, Z coordinate triplets, a list of triangular faces connecting these points and a RGB triplet corresponding to the texture image at each vertex. This dataset is one of the largest 3D facial dataset that is publicly available with 500 different subjects.

4.2.2 German Dataset (200 subjects)

This dataset is provided by Prof. Nikolaus F. Troje from Queens university Canada. It consists for 3D facial meshes of 200 subjects with male to female
ratio of one. All subjects have neutral expression.

4.2.3 BU3DFE dataset (100 subjects)

For expression classification we use BU3D-FE expression dataset [139]. This dataset contains 100 subjects with six different expressions plus neutral face. Each expression except neutral has four intensity levels. Figure 4.2 shows the different expressions intensity levels for a subject from the dataset. Each face has been labeled after voting of multiple subjects which includes experts.
### 4.3 Preprocessing

The raw data is not normalized with respect to scale and orientation. In order to perform a meaningful comparison across subjects, it is essential that we re-orient the faces to a common coordinate system and also account for scale variations. We do this in two steps: pose normalization and scale normalization.
4.3.1 Normalization

For pose normalization we manually select three fiducial points. The three points (figure 4.3.1) are:

a) Left inner canthus (inner corner) of the eye.
b) Right inner canthus of the eye
c) Junction of the nose

The choice of these points is motivated by the fact that these points are relatively stable under facial surface deformation due to expressions etc. Together these points form a plane. The face is rotated such that this plane makes an angle of 30 degrees with X-Y axis. The origin is then shifted to the center of the inner canthus of the eyes. For scale normalization, we apply an affine transformation so that the position of these three fiducial points is same for all the faces. Figure 4.3.1 shows a 3D face before and after normalization. The triangle denotes the plane formed by the fiducial points.
4.4 Feature Extraction

To study the dynamics of the 3D face, it is desired that we extract features which effectively capture the variations. In this study we extract a variety of features including local and global features. Local features such as surface height, orientation of surface normal and curvature capture local variations,
while multi-scale representation, moments etc. capture global variations. The next section describes the features in more details.

4.4.1 Local Feature Extraction

The next subsection describes the local feature extraction process.

4.4.1.1 Local surface descriptors

The original mesh is a a range map which is the distance of each vertex on the mesh from the laser scanner. Besides using this height feature we also use features based on surface normal and curvature. Besides describing a surface in more detail, curvature has the valuable characteristic of being viewpoint invariant. Mathematically, curvature of a surface is the inverse of the radius of a circle that is tangent to that surface at a particular location [57]. For a given point on the surface, there are infinite numbers of circles in the vertical plane with different orientations which are tangent to the surface. The curvature of the circle with the smallest radius is called the maximum curvature (X) and the curvature of the circle with the largest radius is called the minimum curvature (N). Together these are called the principal curvatures. The mean of the principal curvatures is called the mean curvature (H) while the product of principal curvature is called the gaussian curvature (K). Various other curvature
measures can be derived from the principal curvatures. One such measure is the Shape Index. It is a continuous indicator of the local morphological structure of the surface. The shape index ranges from -1 (basin) to 1 (dome). A planer surface has zero shape index. Another surface measure, Curvedness, is a measure of how much curved a surface is. Tables 4.2 and 4.3 shows some of the features mapped on the 3D face. Apart from Orientation (O), direction of maximum (U(X)) and direction minimum curvature (U(N)), all the other features are scalar. O, U(X) and U(N) are 3-element vector features indicating direction in 3D space. For a formal definition and explanation of the different kinds of surface curvatures, the reader is referred to [11, 57]. Equations 4.1 to 4.6 list the mathematical description of some of the features derived from the principal curvatures that we use in this study. We follow [5] for calculation of curvature.

\[ \text{MeanCurvature}(H) = \frac{X + N}{2} \]  
(4.1)

\[ \text{GaussianCurvature}(K) = XN \]  
(4.2)

\[ \text{TotalCurvature}(T) = \frac{|X| + |N|}{2} \]  
(4.3)

\[ \text{RangeCurvature}(R) = \frac{|X| - |N|}{2} \]  
(4.4)

\[ \text{Curvedness}(C) = \frac{2}{\pi} \ln \sqrt{\frac{X^2 + N^2}{2}} \]  
(4.5)
\[ \text{ShapeIndex}(S) = \frac{2}{\pi} \arctan \left( \frac{X + N}{X - N} \right) \] (4.6)

4.4.1.2 Asymmetry of Local surface descriptors

Symmetry can be defined in terms of three transformations in N-Dimensional Euclidean space, \( \mathbb{E}_n \): reflection, rotation and translation. Formally, a subset \( S \) of \( \mathbb{E}_n \) is symmetric with respect to a transformation \( T \) if \( T(S) = S \). Reflection symmetry is a kind of symmetry without rotation, for which one half part is a mirror of the other half part. Reflection symmetry is also referred to as bi-lateral symmetry. The axis of reflection is referred to as the \textit{Symmetry Plane}. The 3D human face shape could be considered as a nearly bi-laterally symmetrical object. (Figure 4.4) For this study we define the symmetry plane as the plane which passes through origin (mid-point of inner canthus of the eyes) and is perpendicular to the plane formed by the three manually selected fiducial points as described in the previous section. Let \( \gamma \) be the bilateral symmetry plane of

Figure 4.4: The Symmetry plane
Differential Geometry based surface features

<table>
<thead>
<tr>
<th>Local Feature</th>
<th>Asymmetry Feature</th>
<th>Local Feature</th>
<th>Asymmetry Feature</th>
</tr>
</thead>
<tbody>
<tr>
<td>Height (Z)</td>
<td></td>
<td>Orientation (O)</td>
<td></td>
</tr>
<tr>
<td>Maximum Curvature (X)</td>
<td>Minimum Curvature (N)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>abs(X)</td>
<td>abs(N)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Direction of (U(X))</td>
<td>Direction (U(N))</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean Curvature (H)</td>
<td>Gaussian Curvature (K) = X*N</td>
<td></td>
<td></td>
</tr>
<tr>
<td>abs(H)</td>
<td>abs(K)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 4.2: The proposed 16 local and asymmetry features mapped to the 3D face of a subject from BJUT-3DR1 dataset. Each feature type has been normalized to [0,1]. The colorbar on the right shows the scale.
Table 4.3: The proposed 16 local and asymmetry features mapped to the 3D face of a subject from BJUT-3DR1 dataset. Each feature type has been normalized to [0,1]. The colorbar on the right shows the scale.

the facial surface. Each point $p_i$ on the facial surface has a corresponding point $p'_i$ when reflected about the plane $\gamma$. If the value of a local feature at a point $p_i$ on the facial surface is $W_i$ then $W'_i$ represents corresponding measure at the reflected point $p'_i$. For a given face $f$ and its corresponding bilateral symmetry plane $\gamma$, we define a local asymmetry score for each of the scalar feature described in the previous as follows:

$$Score(x, y) = |(W_i - W'_i)|$$  (4.7)
Figure 4.5: Decomposition of 3D face to different primitive surface types, e.g. plane, saddle, dome, synform etc based of surface curvature. The variation of surface asymmetry with respect to the symmetry axis under different expressions is clearly seen.

For vector features $O$, $U(X)$ and $U(N)$, the asymmetry score is defined as:

$$Score(x, y) = |(1 - \cos(\beta))|$$  \hspace{1cm} (4.8)

where $\beta$ is the angle between the feature vector at point $p_i$ and $p'_i$. Note that for asymmetry features any half of the face contains all the information, so the dimension of the asymmetry features is half when compared to the local features. Figure 4.5 shows the importance of using asymmetry and curvature features for expression classification. The movement of facial muscles is not symmetric under various expression. Asymmetry features can accurately capture the asymmetry of expressions. The curvature features capture the subtle movements of facial surface as seen in figure 4.5.
4.4.1.3 ICP residual error as a descriptor

Interactive Closest Point (ICP) algorithm is often used to register two point sets when the orientation and scale of the two point sets is roughly the same. ICP iteratively updates a transformation (Translation/Rotation) needed to minimize the distance between two point sets. The iteration continues till a maximum number of iteration is reached or a minimum threshold error is achieved. The output of ICP algorithm is point correspondences, a mean-square error for each correspondence and a transformation to align the two points sets. More details about the ICP algorithm can be found in [144, 109].

The residual MSE error of the ICP algorithm can be used as a feature by registering all the faces to a common neutral face. We use this feature when the relative movement of facial parts is large e.g. expression changes. Figure 4.6 shows the MSE when different level 1 expressions of a subject from BU3DFE was registered to a neutral face. Notice that even small variations in different expressions are easily captured.

4.4.1.4 Statistical of Local Features

For each of facial features described in the previous sections, we calculate a set of 11 statistical features listed in 4.4 in a 5x5 sliding neighborhood. Figure 4.7 shows the visualization of these statistical features for the feature Shape Index.
Figure 4.6: The ICP Mean Square Error as a feature. The figure shows the mean square error of registering various Level 1 expressions for a subject from BU3DFE to a neutral face.

<table>
<thead>
<tr>
<th>Mean</th>
<th>Median</th>
<th>Maximum</th>
<th>Minimum</th>
<th>Mode</th>
</tr>
</thead>
<tbody>
<tr>
<td>Standard Deviation</td>
<td>Entropy</td>
<td>Moment 3,4,5,6</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 4.4: A set of 11 statistical features. See Appendix I for more details.

4.4.2 Grid-Sampling and Multi-Scale representation

The original facial meshes are not evenly sampled, so the number of vertices varies from face to face. The number of vertices is also very large (more than 100,000 for some meshes). Thus, for computational convenience, as well as for the purpose of comparison across datasets, we resample the mesh on a regular cylindrical grid which approximately covers the whole face from chin to forehead in the vertical direction, and from outside outer corner of one eye to the other.
Figure 4.7: Different statistical features for Shape Index for a subject from BU3DFE database.

eye in the horizontal direction. All the features which are calculated on the original mesh are then interpolated on this regular grid via cubic interpolation. Grid size is a parameter which can be varied. If we choose a higher grid size, then we would get many features which are co-related and do not necessarily improve classification. On the other hand a low grid size would give a very crude estimate of the feature. In this study we use a variable grid size of 101x101, 51x51, 25x25 and 13x13 to get multi-resolution estimate of the features. Figure 4.9 shows a multi-resolution approximation of Shape Index for a particular subject. Figure 4.8 shows the Gaussian Curvature of a subject interpolated on a cylindrical grid of 51x51.
4.4.3 Global Feature Extraction

The features described so far capture local variations only. However to capture global variations, we calculate a compact representation by calculating 20 statistical moments and GLCM features (as described in chapter 2) for each feature. We also calculate 256 bin histogram for each feature.
4.5 Results and Discussions

We apply the discrimination pipeline to a number for face discrimination problems. Table 4.5 lists the datasets that were used for various discrimination tasks.

4.5.1 Gender Recognition

As described in table 4.5, we use the BJUT-3DR1 and German dataset for gender classification. The classification was done individually as well as on the combined dataset. The following 8 types of discrimination was done:

1. Chinese Male / Chinese Female
2. German Male / German Female
3. Male (all) / Female (all)
4. Chinese / German
5. Chinese Female / German Female
6. Chinese Male / German Male
7. Chinese Female / German Male
8. Chinese Male / German Female

At each iteration the dataset was divided randomly into training and test set with a ratio of 6:4. The Male-Female ratio was same in both test and training.
<table>
<thead>
<tr>
<th>Discrimination Type</th>
<th>Datasets used</th>
<th>Subject Distribution</th>
<th>Feature Types used</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gender</td>
<td>BU3DFE / German</td>
<td>Chinese: 500 Male/500 Female, German: 97 Male/98 Female</td>
<td>Local and Asymmetry feature only</td>
</tr>
<tr>
<td>Age (classification)</td>
<td>BJUT-3DR1</td>
<td>Chinese: 419 10-29 years/81 30-49 years</td>
<td>Local and Asymmetry feature only</td>
</tr>
<tr>
<td>Age (Regression)</td>
<td>German</td>
<td>continuous</td>
<td>Local and Asymmetry feature only</td>
</tr>
<tr>
<td>Expression</td>
<td>Bu3DFE</td>
<td>100 subjects, 6 expressions, 4 intensity levels</td>
<td>Local/Asymmetry feature + Local statistics + Global features + ICP + RGB</td>
</tr>
<tr>
<td>Gender classification under expression variations</td>
<td>Bu3DFE 100 subjects, 60 Female, 40 Male, 25 scans per subject</td>
<td>Local/Asymmetry features only</td>
<td></td>
</tr>
<tr>
<td>Face Identification under expression variations</td>
<td>Bu3DFE 100 subjects, 60 Female, 40 Male, 25 scans per subject</td>
<td>Local/Asymmetry feature + RGB</td>
<td></td>
</tr>
</tbody>
</table>

Table 4.5: Description of datasets used for various 3D face discrimination experiments

The process was repeated 100 times. The classification results are shown in Table 4.6. Only results for Chinese Male/Female, German Male/Female and Combined Male/Female are shown as for other classes 100 ± 0.0 % classification accuracy was achieved. The results are averaged over 100 iterations. It can be seen that texture plays an important role in gender classification, as without texture, classification rates drops by 3-4% . Another important observation that can be made from the table is the importance for asymmetry features. Adding asymmetry features to the local features, resulted in a significant in-
Table 4.6: Average classification after 100 iterations for Classification of Chinese Male/Female, German Male/Female and Combined Male/Female. Classification rates for other categories was found to be 100%. The number of subjects in Chinese dataset was 500 while 195 subjects from German dataset were used. The best classification rate reported for gender classification is 91% by Jain et al [84] using both texture and range features.

<table>
<thead>
<tr>
<th>Feature Type</th>
<th>CM-CF</th>
<th>GM-GF</th>
<th>M-F</th>
</tr>
</thead>
<tbody>
<tr>
<td>All Features</td>
<td>97.3(±3.4)</td>
<td>94.6(±4.5)</td>
<td>96.3(±3.5)</td>
</tr>
<tr>
<td>All Features without texture</td>
<td>95.9(±2.1)</td>
<td>91.11(±3.4)</td>
<td>92.9(±2.7)</td>
</tr>
<tr>
<td>Local Features</td>
<td>96.2(±2.7)</td>
<td>87.6(±3.5)</td>
<td>94.7(±3.5)</td>
</tr>
<tr>
<td>Local Features without texture</td>
<td>94.5(±2.5)</td>
<td>86.8(±4.1)</td>
<td>90.1(±3.2)</td>
</tr>
<tr>
<td>Asymmetry Features</td>
<td>95.1(±2.5)</td>
<td>84.3(±3.6)</td>
<td>88.7(±2.8)</td>
</tr>
<tr>
<td>Asymmetry Features without texture</td>
<td>92.8(±2.7)</td>
<td>81.8(±3.2)</td>
<td>87.7(±3.1)</td>
</tr>
</tbody>
</table>

crease (p-Value < 0.001) in all the three categories shown in table. Above 80% classification rate was achieved by using asymmetry features alone. This signifies the importance of facial asymmetry for gender classification. Tables 4.8, 4.10 and 4.9 shows the histogram of features selected after AVR, mRMR and feature subset selection stage for some discrimination tasks. Exhaustive results can be found in the supplementary material.
<table>
<thead>
<tr>
<th>S.No</th>
<th>Sym</th>
<th>Name and Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Z</td>
<td>Height</td>
</tr>
<tr>
<td>2</td>
<td>O</td>
<td>Orientation</td>
</tr>
<tr>
<td>3</td>
<td>X</td>
<td>Maximum Curvature</td>
</tr>
<tr>
<td>4</td>
<td>A(X)</td>
<td>Absolute value of maximum curvature</td>
</tr>
<tr>
<td>5</td>
<td>U(X)</td>
<td>Direction of maximum curvature</td>
</tr>
<tr>
<td>6</td>
<td>N</td>
<td>Minimum Curvature</td>
</tr>
<tr>
<td>7</td>
<td>A(N)</td>
<td>Absolute value of minimum curvature</td>
</tr>
<tr>
<td>8</td>
<td>U(N)</td>
<td>Direction of minimum curvature</td>
</tr>
<tr>
<td>9</td>
<td>H</td>
<td>Mean Curvature</td>
</tr>
<tr>
<td>10</td>
<td>A(H)</td>
<td>Absolute value of mean curvature</td>
</tr>
<tr>
<td>11</td>
<td>K</td>
<td>Gaussian Curvature</td>
</tr>
<tr>
<td>12</td>
<td>A(K)</td>
<td>Absolute value of Gaussian curvature</td>
</tr>
<tr>
<td>13</td>
<td>T</td>
<td>Total Curvature</td>
</tr>
<tr>
<td>14</td>
<td>R</td>
<td>Range Curvature</td>
</tr>
<tr>
<td>15</td>
<td>S</td>
<td>Shape Index</td>
</tr>
<tr>
<td>16</td>
<td>C</td>
<td>Curvedness</td>
</tr>
</tbody>
</table>

Table 4.7: List of symbols for 3D facial surface features Note that Orientation (O), Direction of maximum curvature U(X) and Direction of minimum curvature U(N) have X, Y and Z components.
Table 4.8: Table showing histogram of features selected after AVR ranking, mRMR ranking and feature selection for Chinese male and female discrimination and German male and female discrimination when all features were used.
Table 4.9: Table showing histogram of features selected after AVR ranking, mRMR ranking and feature selection for Chinese male and female discrimination when only asymmetry shape and local shape features were used.
Table 4.10: Table showing histogram of features selected after AVR ranking, mRMR ranking and feature selection for German male and female discrimination when only asymmetry shape and local shape features were used.
4.5.1.1 Comparison of Most Discriminating facial regions between male and female

While the previous section described the classification rates for various gender classification tasks, in this section we try to identify regions of the face which are important for gender discrimination. Tables 4.11, 4.12, 4.13, 4.14 show the discriminative scores plotted on 3D face based on AVR ranking of features. The p-Value significance corresponding to table 4.11 and 4.14 are shown in tables 4.15 and 4.16 respectively. Results are shown for only selected feature types only, for exhaustive results reader is referred to the supplementary material. Figure 4.10 shows the combined p-Value and AVR plots for the feature type abs(Minimum curvature). It can be seen that a high value of AVR always has low p-Value. This is further illustrated by figure 4.11.

4.5.1.2 Importance of facial asymmetry

To compare facial asymmetry between different groups, we first find the mean of the top n% of the most discriminative regions of the face based on AVR scores for each subject. Thus we have one mean value for each subject. We then find the p-Value of this score across all subjects in the dataset. Table 4.17 shows the top 1, 10, 50 and 100% of the most discriminating regions which were used to calculate the p-Value for for gender discrimination based on asymmetry features.
Table 4.11: Table showing the most discriminating regions of the face for gender discrimination based on AVR scores. Different columns show discriminating between various classes. Only Height(Z), Orientation Y component (Oy), Mean Curvature (H), absolute value of mean curvature (abs(H)) and Shape Index (S) are shown here.
Table 4.12: Table showing the most discriminating regions of the face for gender discrimination based on AVR scores. Different columns show discriminating between various classes. Only Height (Z), Orientation Y component (Oy), Mean Curvature (H), absolute value of mean curvature (abs(H)) and Shape Index (S) are shown here.

<table>
<thead>
<tr>
<th>F</th>
<th>Chinese Male/German Male</th>
<th>Chinese Female/German Female</th>
<th>Chinese Male/German Female</th>
<th>Chinese Female/German Male</th>
</tr>
</thead>
<tbody>
<tr>
<td>Z</td>
<td><img src="image1" alt="Image" /></td>
<td><img src="image2" alt="Image" /></td>
<td><img src="image3" alt="Image" /></td>
<td><img src="image4" alt="Image" /></td>
</tr>
<tr>
<td>Oy</td>
<td><img src="image5" alt="Image" /></td>
<td><img src="image6" alt="Image" /></td>
<td><img src="image7" alt="Image" /></td>
<td><img src="image8" alt="Image" /></td>
</tr>
<tr>
<td>H</td>
<td><img src="image9" alt="Image" /></td>
<td><img src="image10" alt="Image" /></td>
<td><img src="image11" alt="Image" /></td>
<td><img src="image12" alt="Image" /></td>
</tr>
<tr>
<td>abs(H)</td>
<td><img src="image13" alt="Image" /></td>
<td><img src="image14" alt="Image" /></td>
<td><img src="image15" alt="Image" /></td>
<td><img src="image16" alt="Image" /></td>
</tr>
<tr>
<td>S</td>
<td><img src="image17" alt="Image" /></td>
<td><img src="image18" alt="Image" /></td>
<td><img src="image19" alt="Image" /></td>
<td><img src="image20" alt="Image" /></td>
</tr>
</tbody>
</table>
Table 4.13: Table showing the most discriminating regions of the face for gender discrimination based on AVR scores. Asymmetry feature types for Only Height ($Z$), Orientation $Y$ component ($O_y$), Mean Curvature ($H$), absolute value of mean curvature ($abs(H)$) and Shape Index ($S$) are shown here.
Table 4.14: Table showing the most discriminating regions of the face for gender discrimination based on AVR scores. Only Asymmetry feature types Height ($Z$), Orientation Y component ($O_y$), Mean Curvature ($H$), absolute value of mean curvature ($\text{abs}(H)$) and Shape Index ($S$) are shown here.
Table 4.15: p-Value scores showing statistically significant regions for gender discrimination corresponding to Table 4.13. p-Value above 0.05 has been made transparent, while p-Value below 0.001 has been saturated to 0.001. Only Height(Z), Orientation Y component (Oy), Mean Curvature (H), absolute value of mean curvature (abs(H)) and Shape Index (S) are shown here.
Table 4.16: p-Value scores showing statistically significant regions for gender discrimination corresponding to Table 4.14. p-Value above 0.05 has been made transparent, while p-Value below 0.001 has been saturated to 0.001. Only Height(Z), Orientation Y component (Oy), Mean Curvature (H), absolute value of mean curvature (abs(H)) and Shape Index (S) are shown here.
Figure 4.10: Figure shows various scores for Chinese male/Female discrimination using abs(Min Curvature) feature type. Comprehensive results for other feature types and discrimination tasks can be accessed in the supplementary material. (a) Top left, the region of the face that was used to extract features, (b) Top right, Plot of AVR score and log of p-Value. The black line indicates a significance of 0.05 while the red line indicates a significance of 0.001. (c) Bottom left, p-Value of individual features, (d) AVR score of individual features only. The p-Value is also shown. The p-Value has been thresholded at $10^{-300}$, values below this are shown as zero. It can clearly be seen there is a significant difference between male and female in terms of facial asymmetry as signified by the p-Value scores. Also note the decrease in p-Value as “n” is decreased. Similar results for Chinese verses German discrimination are shown in table 4.18. The box plots of feature distribution in these regions are shown in Table 4.21,
Figure 4.11: Plot of p-Value vs AVR score for feature type abs(Minimum curvature). Top: Linear scale, Bottom: Log scale. It can be seen that AVR score above 0.05 is statistically significant with p-Value <0.001.
4.22, 4.20 and 4.19 for some selected features. Exhaustive results can be found in the supplementary materials. Table 4.21 and 4.19 show feature distribution of subjects for top 1% most discriminative regions for Local and Asymmetry features. Table 4.22 and 4.20 show feature distribution of subjects for top 100% most discriminative regions for Local and Asymmetry features. It can be seen that the difference in facial asymmetry of gender is more in germans than in chinese. Also females are more symmetrical in terms of Height, Orientation and mean curvature in both german and chinese subjects.
<table>
<thead>
<tr>
<th>Feature</th>
<th>Top 1%</th>
<th>Top 10%</th>
<th>Top 50%</th>
<th>Top 100%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Z</td>
<td>![Image]</td>
<td>![Image]</td>
<td>![Image]</td>
<td>![Image]</td>
</tr>
<tr>
<td>p-Value</td>
<td>0</td>
<td>0.37</td>
<td>0.94</td>
<td>0.13</td>
</tr>
<tr>
<td>Oy</td>
<td>![Image]</td>
<td>![Image]</td>
<td>![Image]</td>
<td>![Image]</td>
</tr>
<tr>
<td>p-Value</td>
<td>0</td>
<td>2.9×10^{-9}</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>H</td>
<td>![Image]</td>
<td>![Image]</td>
<td>![Image]</td>
<td>![Image]</td>
</tr>
<tr>
<td>p-Value</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>abs(H)</td>
<td>![Image]</td>
<td>![Image]</td>
<td>![Image]</td>
<td>![Image]</td>
</tr>
<tr>
<td>p-Value</td>
<td>1.24×10^{-5}</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>S</td>
<td>![Image]</td>
<td>![Image]</td>
<td>![Image]</td>
<td>![Image]</td>
</tr>
<tr>
<td>p-Value</td>
<td>0</td>
<td>0</td>
<td>6.9×10^{-14}</td>
<td>1.8×10^{-11}</td>
</tr>
</tbody>
</table>

Table 4.17: Table showing the most discriminating asymmetric regions of the face for Male versus Female discrimination for some feature types. The p-Value is also shown. The green region indicates the region from where sample means was taken. Only Height(Z), Orientation Y component (Oy), Mean Curvature (H), absolute value of mean curvature (abs(H)) and Shape Index (S) are shown here.
Table 4.18: Table showing the most discriminating asymmetric regions of the face for German verses Chinese discrimination for some feature types. The p-Value is also shown. The p-Value is also shown. The green region indicates the region from where sample means was taken. Only Height(Z), Orientation Y component (Oy), Mean Curvature (H), absolute value of mean curvature (abs(H)) and Shape Index (S) are shown here.
Table 4.19: Table showing histogram of mean values for asymmetry features for various discrimination tasks when top 1% of the features was taken into account. Only Height(Z), Asymmetry Orientation(Na), Mean Curvature (H), Gaussian Curvature (K) and Shape Index (S) are shown here.
Table 4.20: Table showing histogram of mean values for asymmetry features for various discrimination tasks when top 1% of the features was taken into account. Only Height(Z), Asymmetry Orientation(Na), Mean Curvature (H), Gaussian Curvature (K) and Shape Index (S) are shown here.
Table 4.21: Table showing histogram of mean values for Local features for various discrimination tasks when top 1% of the features was taken into account. Only Height(Z), Asymmetry Orientation(Na), Mean Curvature (H), Gaussian Curvature (K) and Shape Index (S) are shown here.
Table 4.22: Table showing histogram of mean values for Local features for various discrimination tasks when top 1% of the features was taken into account. Only Height(Z), Asymmetry Orientation(Na), Mean Curvature (H), Gaussian Curvature (K) and Shape Index (S) are shown here.
<table>
<thead>
<tr>
<th></th>
<th>Sensitivity (Class 1)</th>
<th>Specificity (Class 2)</th>
<th>Overall</th>
</tr>
</thead>
<tbody>
<tr>
<td>Training</td>
<td>100.0( ± 0.0)</td>
<td>100.0( ± 0.0)</td>
<td>100.0( ± 0.0)</td>
</tr>
<tr>
<td>Test</td>
<td>91.8( ± 5.4)</td>
<td>73.7( ± 6.6)</td>
<td>88.8( ± 5.7)</td>
</tr>
</tbody>
</table>

Table 4.23: Age classification results on the Chinese dataset

4.5.2 Age Estimation

For age estimation for facial mesh, we perform two sets of experiments. For the first experiment we perform age classification on the BJUT3DR1 dataset, for the second we perform age regression on the German dataset.

4.5.2.1 Age Classification

The BJUT-3DR1 Chinese dataset has subjects of age ranging from 16 to 49. Only the 10 year period age label is known for the subjects e.g. 10-19, 20-29, 30, 39 and 40-49. As a lot of subjects lie in the 20-29 age group, we further grouped subjects in 10-29 in class 1 and 30-49 age group in class 2. Only local surface features were used for this experiment. We split the test and training in the ratio of 4:6. Feature ranking was done using AVR scores followed by mRMR redundancy removal and feature subset selection. The process was repeated 100 times. The results of classification are shown in table 4.23. Table 4.24 shows the discriminating regions of the face for BJUT3DR1 dataset based on AVR scores.
<table>
<thead>
<tr>
<th>Local Features</th>
<th>Asymmetry Features</th>
</tr>
</thead>
<tbody>
<tr>
<td>Height</td>
<td>Height</td>
</tr>
<tr>
<td>NorX</td>
<td>NorX</td>
</tr>
<tr>
<td>Min Curvature</td>
<td>Min Curvature</td>
</tr>
<tr>
<td>Max Curvature</td>
<td>Max Curvature</td>
</tr>
<tr>
<td>Gauss Curvature</td>
<td>Gauss Curvature</td>
</tr>
<tr>
<td>Mean Curvature</td>
<td>Mean Curvature</td>
</tr>
<tr>
<td>Total Curvature</td>
<td>Total Curvature</td>
</tr>
<tr>
<td>Range Curvature</td>
<td>Range Curvature</td>
</tr>
<tr>
<td>Curvedness</td>
<td>Curvedness</td>
</tr>
<tr>
<td>Shape Index</td>
<td>Shape Index</td>
</tr>
</tbody>
</table>

Table 4.24: Table showing the most discriminating regions of the face for age discrimination based on AVR scores on the CHinese dataset. Asymmetry feature types are also shown.
4.5.2.2 Age Regression

The German dataset has age label of subjects in months. The age varies from 206 months (17.16 years) to 549 months (45.75 years). We divide the dataset into 140 subjects for training and 48 subjects for test. Mutual Information was used for feature ranking followed by mRMR redundancy removal and forward feature selection. The process was repeated 100 times. A mean training error of 2.4 ± 1.9 years and a mean test error of 4.0 ± 3.4 years was achieved. Figure 4.12 shows the age prediction and ground truth for test and training set. The absolute error is shown in figure 4.13. Table 4.25 and 4.26 show the most discriminating regions of the face for a subset of features based on mutual information and correlation respectively.
Table 4.25: Table showing the most discriminating regions of the face for age regression based on Mutual Information scores on the German dataset. Asymmetry feature types are also shown.
<table>
<thead>
<tr>
<th>Local Features</th>
<th>Asymmetry Features</th>
</tr>
</thead>
<tbody>
<tr>
<td>Height</td>
<td>Height</td>
</tr>
<tr>
<td></td>
<td>NorX</td>
</tr>
<tr>
<td>Min Curvature</td>
<td>Min Curvature</td>
</tr>
<tr>
<td></td>
<td>Max Curvature</td>
</tr>
<tr>
<td>Gauss Curvature</td>
<td>Gauss Curvature</td>
</tr>
<tr>
<td></td>
<td>Mean Curvature</td>
</tr>
<tr>
<td>Total Curvature</td>
<td>Total Curvature</td>
</tr>
<tr>
<td></td>
<td>Range Curvature</td>
</tr>
<tr>
<td>Curvedness</td>
<td>Curvedness</td>
</tr>
<tr>
<td></td>
<td>Shape Index</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 4.26: Table showing the most discriminating regions of the face for age discrimination based on Correlation scores on the German dataset. Asymmetry feature types are also shown.
Figure 4.12: Age prediction on Test and Training set
Figure 4.13: Absolute Age prediction error on Test and Training set
4.5.3 Expression Classification

The BU3DFE dataset has subjects with 7 different expressions namely Anger, Disgust, Surprise, Sad, Happy, Fear and Neutral. Except for neutral expression, all the other types have 4 intensity levels. 90 subjects are used for training and rest used for testing. Table 4 shows the average confusion matrix for Level 4 expression classification after 100 iterations. We see that anger and sadness are most difficult to classify and are often classified as neutral. This can be attributed to the fact that there is relatively less movement of the face when compared to other expressions. For example, surprise generally includes large movement of the mouth and can be easily distinguished form other expressions. This is reflected in its classification rate which is the best among all expressions. Figure 4.14 shows the distribution of subjects in top 3 mRMR reduced dimensions for one random split of training and test set. It is seen that Happy, surprise and disgust form a relatively distinct cluster, while all the other expression types are relatively overlapping. This is also reflected in the confusion matrix where the classification rates of the former is in the 90s while that of the later is in the 80s. A best mean overall classification rate of $88.87\pm3.15\%$ is achieved using a LDA classifier. The confusion matrix for level 3, level 2 and level 1 are shown in table 4.28, 4.29 and 4.30 respectively. A classification rate of $78.49\%$, $73.59\%$ and $67.71\%$ is achieved for expression classification of level
Table 4.27: Confusion matrix showing expression classification rates and standard deviation averaged over 100 iterations on the BU3DFE dataset. The average overall classification rates is 88.87±3.15%. The best classification rate reported on this dataset is by Tang and Huang [116], who achieved 95.1% by using features extracted from 83 manually labelled fiducial points.

<table>
<thead>
<tr>
<th></th>
<th>Anger</th>
<th>Disgust</th>
<th>Fear</th>
<th>Happy</th>
<th>Sad</th>
<th>Surprise</th>
<th>Neutral</th>
</tr>
</thead>
<tbody>
<tr>
<td>AN</td>
<td>82.5 ± 2.4</td>
<td>2.5 ± 0.0</td>
<td>3.7 ± 0.0</td>
<td>0.0 ± 0.0</td>
<td>3.7 ± 0.0</td>
<td>1.2 ± 0.0</td>
<td>6.2 ± 0.0</td>
</tr>
<tr>
<td>DI</td>
<td>1.3 ± 0.0</td>
<td>93.5 ± 2.7</td>
<td>3.9 ± 0.0</td>
<td>1.3 ± 0.0</td>
<td>0.0 ± 0.0</td>
<td>0.0 ± 0.0</td>
<td>0.0 ± 0.0</td>
</tr>
<tr>
<td>FE</td>
<td>2.5 ± 0.0</td>
<td>5.0 ± 0.0</td>
<td>83.5 ± 3.6</td>
<td>1.2 ± 0.0</td>
<td>2.5 ± 0.0</td>
<td>0.0 ± 0.0</td>
<td>5.0 ± 0.0</td>
</tr>
<tr>
<td>HA</td>
<td>0.0 ± 0.0</td>
<td>1.2 ± 0.0</td>
<td>5.1 ± 0.0</td>
<td>93.5 ± 3.3</td>
<td>0.0 ± 0.0</td>
<td>0.0 ± 0.0</td>
<td>0.0 ± 0.0</td>
</tr>
<tr>
<td>SA</td>
<td>5.0 ± 0.0</td>
<td>0.0 ± 0.0</td>
<td>1.2 ± 0.0</td>
<td>0.0 ± 0.0</td>
<td>82.2 ± 2.3</td>
<td>0.0 ± 0.0</td>
<td>11.3 ± 0.1</td>
</tr>
<tr>
<td>SU</td>
<td>0.0 ± 0.0</td>
<td>1.5 ± 0.0</td>
<td>0.0 ± 0.0</td>
<td>0.0 ± 0.0</td>
<td>0.0 ± 0.0</td>
<td>96.9 ± 6.2</td>
<td>1.5 ± 0.0</td>
</tr>
<tr>
<td>NE</td>
<td>3.8 ± 0.0</td>
<td>0.0 ± 0.0</td>
<td>5.0 ± 0.0</td>
<td>0.0 ± 0.0</td>
<td>1.2 ± 0.0</td>
<td>0.0 ± 0.0</td>
<td>89.8 ± 1.6</td>
</tr>
</tbody>
</table>

Table 4.28: Level 3 Confusion matrix showing mean expression classification rates averaged over 10 iterations on the BU3DFE dataset. The average overall classification rates is 78.49%

<table>
<thead>
<tr>
<th></th>
<th>Anger</th>
<th>Disgust</th>
<th>Fear</th>
<th>Happy</th>
<th>Sad</th>
<th>Surprise</th>
<th>Neutral</th>
</tr>
</thead>
<tbody>
<tr>
<td>AN</td>
<td>63.75</td>
<td>5.00</td>
<td>3.75</td>
<td>1.25</td>
<td>16.25</td>
<td>0</td>
<td>10.00</td>
</tr>
<tr>
<td>DI</td>
<td>5.26</td>
<td>73.68</td>
<td>10.53</td>
<td>3.95</td>
<td>2.63</td>
<td>3.95</td>
<td>0</td>
</tr>
<tr>
<td>FE</td>
<td>3.90</td>
<td>7.79</td>
<td>67.53</td>
<td>10.39</td>
<td>2.60</td>
<td>3.90</td>
<td>3.90</td>
</tr>
<tr>
<td>HA</td>
<td>0</td>
<td>1.27</td>
<td>3.80</td>
<td>94.94</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>SA</td>
<td>10.13</td>
<td>1.27</td>
<td>1.27</td>
<td>0</td>
<td>73.42</td>
<td>0</td>
<td>13.92</td>
</tr>
<tr>
<td>SU</td>
<td>0</td>
<td>1.45</td>
<td>2.90</td>
<td>1.45</td>
<td>0</td>
<td>91.30</td>
<td>2.90</td>
</tr>
<tr>
<td>NE</td>
<td>13.92</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1.27</td>
<td>0</td>
<td>84.81</td>
</tr>
</tbody>
</table>

3, 2 and 1 respectively. Table 4.31 and 4.32 show the confusion matrix for level 4 classification when only local and asymmetry features were used respectively.
<table>
<thead>
<tr>
<th>Level 2:</th>
<th>Anger</th>
<th>Disgust</th>
<th>Fear</th>
<th>Happy</th>
<th>Sad</th>
<th>Surprise</th>
<th>Neutral</th>
</tr>
</thead>
<tbody>
<tr>
<td>65.00</td>
<td>6.33</td>
<td>2.53</td>
<td>0</td>
<td>3.80</td>
<td>18.99</td>
<td>0</td>
<td>6.33</td>
</tr>
<tr>
<td>70.89</td>
<td>10.13</td>
<td>2.63</td>
<td>0</td>
<td>1.27</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>0</td>
<td>3.80</td>
<td>5.26</td>
<td>1.27</td>
<td>0</td>
<td>11.39</td>
<td>0</td>
<td>81.01</td>
</tr>
<tr>
<td>15.00</td>
<td>3.80</td>
<td>5.06</td>
<td>8.86</td>
<td>1.27</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>3.80</td>
<td>5.06</td>
<td>6.33</td>
<td>1.27</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>6.33</td>
<td>5.06</td>
<td>8.86</td>
<td>6.33</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>15.00</td>
<td>3.80</td>
<td>5.06</td>
<td>6.33</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

Table 4.29: Level 2 Confusion matrix showing mean expression classification rates averaged over 10 iterations on the BU3DFE dataset. The average overall classification rates is 73.59%.

<table>
<thead>
<tr>
<th>Level 1:</th>
<th>Anger</th>
<th>Disgust</th>
<th>Fear</th>
<th>Happy</th>
<th>Sad</th>
<th>Surprise</th>
<th>Neutral</th>
</tr>
</thead>
<tbody>
<tr>
<td>65.00</td>
<td>7.59</td>
<td>5.06</td>
<td>2.53</td>
<td>0</td>
<td>20.25</td>
<td>0</td>
<td>17.72</td>
</tr>
<tr>
<td>3.75</td>
<td>70.89</td>
<td>17.72</td>
<td>3.80</td>
<td>0</td>
<td>5.41</td>
<td>2.70</td>
<td>2.53</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>2.53</td>
<td>10.13</td>
<td>3.80</td>
<td>12.70</td>
<td>0</td>
<td>2.53</td>
</tr>
<tr>
<td>18.75</td>
<td>5.06</td>
<td>8.86</td>
<td>6.33</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>17.72</td>
<td>77.22</td>
<td>8.86</td>
<td>2.53</td>
<td>0</td>
<td>82.50</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>12.70</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>68.34</td>
<td>0</td>
<td>2.70</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

Table 4.30: Level 1 Confusion matrix showing mean expression classification rates averaged over 10 iterations on the BU3DFE dataset. The average overall classification rates is 67.71%.

<table>
<thead>
<tr>
<th></th>
<th>Anger</th>
<th>Disgust</th>
<th>Fear</th>
<th>Happy</th>
<th>Sad</th>
<th>Surprise</th>
<th>Neutral</th>
</tr>
</thead>
<tbody>
<tr>
<td>Anger</td>
<td>74.6700</td>
<td>5.3300</td>
<td>2.6700</td>
<td>0</td>
<td>8.0000</td>
<td>0</td>
<td>9.3300</td>
</tr>
<tr>
<td>Disgust</td>
<td>1.3500</td>
<td>87.8400</td>
<td>5.4100</td>
<td>1.3500</td>
<td>0</td>
<td>1.3500</td>
<td>2.7000</td>
</tr>
<tr>
<td>Fear</td>
<td>1.3500</td>
<td>2.7000</td>
<td>75.6800</td>
<td>6.7600</td>
<td>5.4100</td>
<td>4.0500</td>
<td>4.0500</td>
</tr>
<tr>
<td>Happy</td>
<td>0</td>
<td>0</td>
<td>5.4100</td>
<td>93.2400</td>
<td>0</td>
<td>1.3500</td>
<td>0</td>
</tr>
<tr>
<td>Sad</td>
<td>5.4100</td>
<td>1.3500</td>
<td>6.7600</td>
<td>0</td>
<td>75.6800</td>
<td>0</td>
<td>10.8100</td>
</tr>
<tr>
<td>Surprise</td>
<td>0</td>
<td>0</td>
<td>2.7000</td>
<td>1.3500</td>
<td>0</td>
<td>95.9500</td>
<td>0</td>
</tr>
<tr>
<td>Neutral</td>
<td>1.3500</td>
<td>2.7000</td>
<td>2.7000</td>
<td>0</td>
<td>8.1100</td>
<td>0</td>
<td>85.1400</td>
</tr>
</tbody>
</table>

Table 4.31: Confusion matrix for the classification rates (over all 84.0286±4.7%) when only full face features were used.
Figure 4.14: Feature Distribution in top 3 mRMR feature space during one random split for expression classification

<table>
<thead>
<tr>
<th></th>
<th>Anger</th>
<th>Disgust</th>
<th>Fear</th>
<th>Happy</th>
<th>Sad</th>
<th>Surprise</th>
<th>Neutral</th>
</tr>
</thead>
<tbody>
<tr>
<td>Anger</td>
<td>62.670</td>
<td>5.3300</td>
<td>4.0000</td>
<td>0</td>
<td>14.670</td>
<td>0</td>
<td>13.3300</td>
</tr>
<tr>
<td>Disgust</td>
<td>5.4100</td>
<td>78.3800</td>
<td>5.4100</td>
<td>2.7000</td>
<td>2.7000</td>
<td>2.7000</td>
<td>2.7000</td>
</tr>
<tr>
<td>Fear</td>
<td>5.4100</td>
<td>5.4100</td>
<td>71.6200</td>
<td>8.1100</td>
<td>4.0500</td>
<td>1.3500</td>
<td>4.0500</td>
</tr>
<tr>
<td>Happy</td>
<td>0</td>
<td>1.3500</td>
<td>6.7600</td>
<td>91.8900</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Sad</td>
<td>9.4600</td>
<td>0</td>
<td>8.1100</td>
<td>1.3500</td>
<td>68.9200</td>
<td>0</td>
<td>12.1600</td>
</tr>
<tr>
<td>Surprise</td>
<td>0</td>
<td>1.3500</td>
<td>0</td>
<td>1.3500</td>
<td>1.3500</td>
<td>91.8900</td>
<td>4.0500</td>
</tr>
<tr>
<td>Neutral</td>
<td>5.4100</td>
<td>0</td>
<td>4.0500</td>
<td>0</td>
<td>2.7000</td>
<td>0</td>
<td>87.8400</td>
</tr>
</tbody>
</table>

Table 4.32: Confusion matrix for the classification rates (over all 79.03±5.8%) when only asymmetry features were used
4.5.4 Expression invariant gender classification

For expression invariant gender classification we use 43 subjects for training and 40 for test. The Male to Female ratio is one for both training and test set. Only level 4 expression level was evaluated. The best classification rate of 81±4.3% is achieved using a LDA classifier. It was found that forehead is more important for discriminating gender under fear, happy, surprise and sad expression type while cheeks are more important for disgust. End of the lips are more important for gender discrimination under anger. Similar results are obtained using different feature types.

In another experiment, gender classification was done separately for various expressions. Table 4.33 summarizes the results. It can be seen that, apart from disgust, there is more than 10% difference between the sensitivity and specificity for gender classification. Also the mean classification rate is around 75%. This clearly indicates that there is a difference between how males and females express themselves facially. Table 4.35 and 4.34 show the discinative features of the face for gender discrimination under various expressions for some feature types.
Table 4.33: Gender classification rates under different expressions

<table>
<thead>
<tr>
<th>Feat</th>
<th>Sensitivity</th>
<th>Specificity</th>
<th>Overall</th>
</tr>
</thead>
<tbody>
<tr>
<td>Angry</td>
<td>90.0±0.1</td>
<td>70.2±8.6</td>
<td>80.0±4.3</td>
</tr>
<tr>
<td>Disgust</td>
<td>70±28.2</td>
<td>75±21.2</td>
<td>72.5±3.5</td>
</tr>
<tr>
<td>Fear</td>
<td>60±28.2</td>
<td>90±14.3</td>
<td>75±7.0</td>
</tr>
<tr>
<td>Happy</td>
<td>70.0±13.7</td>
<td>55.1±21.2</td>
<td>62.5±2.9</td>
</tr>
<tr>
<td>Sad</td>
<td>65.0±7.0</td>
<td>85.0±6.1</td>
<td>75.0±7.0</td>
</tr>
<tr>
<td>Surprise</td>
<td>90.0±7.0</td>
<td>75.0±16.0</td>
<td>82±14.1</td>
</tr>
<tr>
<td>Neutral</td>
<td>85.3±6.0</td>
<td>90.0±8.0</td>
<td>87.3±5.1</td>
</tr>
<tr>
<td>F</td>
<td>Height</td>
<td>Mean Curvature</td>
<td>Gaussian Curvature</td>
</tr>
<tr>
<td>----------</td>
<td>--------</td>
<td>----------------</td>
<td>-------------------</td>
</tr>
<tr>
<td>Anger</td>
<td><img src="image1" alt="Image" /></td>
<td><img src="image2" alt="Image" /></td>
<td><img src="image3" alt="Image" /></td>
</tr>
<tr>
<td>Disgust</td>
<td><img src="image5" alt="Image" /></td>
<td><img src="image6" alt="Image" /></td>
<td><img src="image7" alt="Image" /></td>
</tr>
<tr>
<td>Surprise</td>
<td><img src="image9" alt="Image" /></td>
<td><img src="image10" alt="Image" /></td>
<td><img src="image11" alt="Image" /></td>
</tr>
<tr>
<td>Sad</td>
<td><img src="image13" alt="Image" /></td>
<td><img src="image14" alt="Image" /></td>
<td><img src="image15" alt="Image" /></td>
</tr>
<tr>
<td>Fear</td>
<td><img src="image17" alt="Image" /></td>
<td><img src="image18" alt="Image" /></td>
<td><img src="image19" alt="Image" /></td>
</tr>
<tr>
<td>Happy</td>
<td><img src="image21" alt="Image" /></td>
<td><img src="image22" alt="Image" /></td>
<td><img src="image23" alt="Image" /></td>
</tr>
<tr>
<td>Neutral</td>
<td><img src="image25" alt="Image" /></td>
<td><img src="image26" alt="Image" /></td>
<td><img src="image27" alt="Image" /></td>
</tr>
</tbody>
</table>

Table 4.34: Discriminative regions for gender identification (local features)
<table>
<thead>
<tr>
<th>F</th>
<th>Height</th>
<th>Mean Curvature</th>
<th>Gaussian Curvature</th>
<th>Shape Index</th>
</tr>
</thead>
<tbody>
<tr>
<td>Anger</td>
<td><img src="image1" alt="Image" /></td>
<td><img src="image2" alt="Image" /></td>
<td><img src="image3" alt="Image" /></td>
<td><img src="image4" alt="Image" /></td>
</tr>
<tr>
<td>Disgust</td>
<td><img src="image5" alt="Image" /></td>
<td><img src="image6" alt="Image" /></td>
<td><img src="image7" alt="Image" /></td>
<td><img src="image8" alt="Image" /></td>
</tr>
<tr>
<td>Surprise</td>
<td><img src="image9" alt="Image" /></td>
<td><img src="image10" alt="Image" /></td>
<td><img src="image11" alt="Image" /></td>
<td><img src="image12" alt="Image" /></td>
</tr>
<tr>
<td>Sad</td>
<td><img src="image13" alt="Image" /></td>
<td><img src="image14" alt="Image" /></td>
<td><img src="image15" alt="Image" /></td>
<td><img src="image16" alt="Image" /></td>
</tr>
<tr>
<td>Fear</td>
<td><img src="image17" alt="Image" /></td>
<td><img src="image18" alt="Image" /></td>
<td><img src="image19" alt="Image" /></td>
<td><img src="image20" alt="Image" /></td>
</tr>
<tr>
<td>Happy</td>
<td><img src="image21" alt="Image" /></td>
<td><img src="image22" alt="Image" /></td>
<td><img src="image23" alt="Image" /></td>
<td><img src="image24" alt="Image" /></td>
</tr>
<tr>
<td>Neutral</td>
<td><img src="image25" alt="Image" /></td>
<td><img src="image26" alt="Image" /></td>
<td><img src="image27" alt="Image" /></td>
<td><img src="image28" alt="Image" /></td>
</tr>
</tbody>
</table>

Table 4.35: Discriminative regions for gender identification (asymmetry features)
Table 4.36: Classification rates ± standard deviation averaged over 100 iterations when using (a)symmetry features alone, local features alone and when both are combined.

<table>
<thead>
<tr>
<th>Task</th>
<th>Asymmetry</th>
<th>Local</th>
<th>Combined</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gender</td>
<td>73.4 ±3.0</td>
<td>80.4 ±3.8</td>
<td>81.0 ±4.3</td>
</tr>
<tr>
<td>Exp</td>
<td>79.2 ±6.2</td>
<td>82.5 ±4.5</td>
<td>88.8 ±3.1</td>
</tr>
<tr>
<td>Face ID</td>
<td>76.1 ±2.3</td>
<td>85.5 ±10.6</td>
<td>87.6 ±5.3</td>
</tr>
</tbody>
</table>

### 4.5.5 Face identification

The BU3DFE dataset contains 100 subjects with 25 different expressions each. These include a neutral face and 4 intensity levels for 6 basic expression types. We consider this as a multi-class classification problem with 100 classes. 20 faces are randomly selected from each class and used for training while the rest 5 faces are used for testing. We achieve a mean classification accuracy of 87.6±5.3% after 100 iterations. Table 4.36 summarizes the results of expression invariant gender classification, expression classification and face identification. It also lists the average classification rates for various discrimination tasks when using asymmetry feature alone, local feature alone and when they are combined. The rates are averaged over 100 iterations. It can be seen that by using asymmetry feature alone, we achieve a mean classification rate in the higher 70s. Specifically for expression classification it is as high as 79.2±6.2%. When combined with local features, a significant (p-Value <0.001) improvement in classification rate is achieved.
Figure 4.15: Relative importance of local height features (101x101 grid) for various discrimination tasks. Note that features that are important for one task may not be for other discrimination task. This is similar to results reported by Liu et al[81], where it was shown that different features were relatively important for face identification and expression recognition for the 2D case. Table 4.37 and 4.38 show some of the features mapped to the 3D face.
<table>
<thead>
<tr>
<th>Feature</th>
<th>Gender</th>
<th>Expression</th>
<th>Face ID</th>
</tr>
</thead>
<tbody>
<tr>
<td>Height</td>
<td>![Height]</td>
<td>![Height]</td>
<td>![Height]</td>
</tr>
<tr>
<td>Mean Curvature</td>
<td>![Mean Curvature]</td>
<td>![Mean Curvature]</td>
<td>![Mean Curvature]</td>
</tr>
<tr>
<td>Gaussian Curvature</td>
<td>![Gaussian Curvature]</td>
<td>![Gaussian Curvature]</td>
<td>![Gaussian Curvature]</td>
</tr>
<tr>
<td>Shape Index</td>
<td>![Shape Index]</td>
<td>![Shape Index]</td>
<td>![Shape Index]</td>
</tr>
<tr>
<td>Curvedness</td>
<td>![Curvedness]</td>
<td>![Curvedness]</td>
<td>![Curvedness]</td>
</tr>
</tbody>
</table>

Table 4.37: Table showing relative importance of local features for various discrimination task on the BU3DFE dataset
Table 4.38: Table showing relative importance of asymmetry features for various discrimination task on the BU3DFE dataset
Chapter 5

Neuroradiology Image Analysis

Men and Women are different. But, aside from external anatomical and primary and secondary sexual differences, is the human brain “sexed” too? What region of the brain is responsible for diseases such as schizophrenia, Alzheimer’s etc? Do these diseases result in some structural changes in the brain? Questions such as these has intrigued researchers from diverse fields. We attempt to answers some of these questions by applying a discriminative analysis on features extracted from a large sample of human brain structural MRI. The next section describes the datasets and feature extraction process.

5.1 3D Structural brain MRI datasets

The dataset used in this study are described below.
5.1.1 Aging Dataset

This dataset consists of structural T1 MR images of 261 subjects. These images were acquired on GE 1.5T Signa scanner between 1999 and 2004 at the University of Pittsburgh Alzheimer’s Research Center. The individuals for the study were selected by neuro-radiologists at the University of Pittsburgh Medical Center. No participant has a neurological disease and all have similar educational level. Mean age of the participants is 66 years (SD=7.5). The subjects age range from 50 to 79 years. Figure 5.1 shows the distribution of subjects in the aging dataset. Both gender and age details are available for all the subjects. There are 129 females and 132 male subjects in the dataset.
Figure 5.2: Cross section of a subject from the aging dataset.

5.1.2 Alzheimer’s Dataset (ADRC)

This dataset has MR images of 66 subjects, each selected by experts from the Alzheimer’s Disease Research Center. There are 20, 26 and 20 subjects in normal controls (CTL), mild cognitive impairment (MCI) and diagnosed probable Alzheimer’s Disease (AD) category respectively. All subjects have comparable age, education and sex. A 1.5 Tesla GE Signa scanner was used to acquire 120 thin, contiguous image slices throughout the brain. The contrast was designed to maximize the gray-white matter and CSF differences.
5.1.3 Cardiovascular Health Study (CHS) Dataset

As part of Cardiovascular Health Study (CHS) Cognition Study at the Pittsburgh center 389 subjects had high-resolution MR scans in 1997-99. This dataset is referred to CHS 3 dataset. From this sample 207 subjects were rescanned in 2002-04. This set is referred to as CHS 1. Out of these 71 subjects received a third scan in 2004-05. This set is called CHS 2 dataset. At the time of the first scan only 2 subjects were diagnosed with mild cognitive impairment (MCI). However during further scans, many of the patients developed dementia or MCI. Thus by using future labels with the original scans and performing discriminative analysis, we can identify regions of the brain which changed due to these diseases. Even though some of the subjects were normal at the time of the scan, by using labels from future diagnosis we try to identify if we can predict neurological disease at an early stage. Note that, the diagnosis of all the patients is not available also some of the images had registration errors, so we use a subset of the subjects mentioned above for our experiments. Table 5.4 shows the details of number of subjects used for CHS 1, CHS 2 and CHS 3 datasets.
5.2 Feature Extraction

The feature extraction framework is divided into 2 stages: 1. Registration 2. Feature Extraction

First we register all the brains to a reference and then extract various features. These features are then used for discriminative analysis and classification. We finally visualize which regions of the brain play an important part in discriminating. The whole process is described in the sections below.

5.3 Deformable Registration framework

The orientation of the subjects head in the original brain scans is not uniform across the whole dataset. Thus to perform any meaningful comparison, all the scans are registered to a standard brain template (MNI152). We do this in two steps, first we affinely register the images, which is followed by dense deformable registration.

5.3.1 Brain Atlas

Talairach [115] and Montreal Neurological Institute (MNI) [29, 28] stereotaxical atlas are the two most widely used brain atlas. The Talairach atlas was created by dissecting and photographing the Talairach brain. However the Talairach
atlas labels are only approximate. The MNI digital brain atlas was created to be more representative of the population. The MNI space was defined by creating average of 305 T1-weighted volumes using an automatic cross-correlation. Later the model was refined by registering scans of 152 brains to the MNI305 average brain. In this study we use the Colin27 brain registered in the MNI152 space.

5.3.2 Brain Atlas Labels

To estimate the approximate position of the brain, the anatomical regions of the MNI brain was labeled in the Automated Anatomical Labeling (AAL) atlas [123]. There are 116 anatomical regions labeled in the AAL atlas. Figure 5.5 shows the AAL atlas labels over the MNI brain image. However there are many regions of the brain which are not labeled in the AAL atlas. In this project we use a fully labeled human brain provided Harvard Medical School. All parts of the brain were labeled on a scan of one subject. This scan was registered to the MNI152 and the labels were transferred in the MNI space. Figure 5.3 shows the Harvard Atlas before and after registration to MNI atlas. The rendering of the labels for AAL and Harvard atlas are shown in 5.6.
5.3.3 Preprocessing

To prepare images for the fully deformable registration, we perform intensity normalization using a histogram equalization followed by mid-sagittal plane (MSP) alignment [76]. The MSP algorithm estimates the bi-lateral symmetry plane and reorients the images. At the last preparatory step, we affinely register all the images in the dataset to the reference image, MNI152. The preliminary steps reduce brightness variations and global differences in overall size, orientation, location and skewing between the images in the dataset. It also increases the convergence time of the subsequent deformable registration. Figure 8 shows the preprocessing pipeline. Note that the final MSP block estimates the deviation from ideal MSP after affine registration. If the registration is good, this deviation should be very less. Thus it acts as an indicator of whether to use or discard a sample.
Figure 5.4: Preprocessing pipeline.
Figure 5.5: Cross section and 3D view of MNI152 brain that was used to register all the MRI scans in this dataset. The bottom row shows the labelled atlas.

5.3.4 3D Deformable registration

A 3D affine transformation is characterized by a coordinate transform defined as a $4 \times 4$ matrix $A$ that is uniquely specified by 12 values (3 translations, 3
rotations, 3 zooms and 3 shears). This is sufficient to model overall differences in position and size between different brains. However it is often not sufficient for modeling more local differences such as e.g. enlarged ventricles or sulci. One may then instead use a non-linear transformation given by equation 5.2.

\[
\begin{bmatrix}
x' \\
y' \\
z' \\
1
\end{bmatrix}
= A \begin{bmatrix}
x \\
y \\
z \\
1
\end{bmatrix}
\]  

(5.1)

The crucial part here are the di(x,y,z), the warp fields. The warp are often represented as linear combinations of basis-functions, thereby reducing the number of parameters that needs to be determined. Some of the issues that have to be taken into consideration during deformable registration are Cost-function, deformation model, optimization criterion, regularization and intensity modulation.
We use a B-Spline, mutual information based deformation model implemented in the insight toolkit. Thus we deformably register all the brains to a standard reference. One of the advantages of using a standard reference such as MNI152 is that, it is widely studied and labelled atlas are present for them. Thus we can automatically identify which voxel corresponds to which part of the brain. Figure 5.5 shows the atlas for MNI152 brain with 169 different parts labelled.

\[ \begin{bmatrix}
    x' \\
    y' \\
    z' \\
    1
\end{bmatrix} = A \begin{bmatrix}
    x \\
    y \\
    z \\
    1
\end{bmatrix} + \begin{bmatrix}
    d_x(x, y, z) \\
    d_y(x, y, z) \\
    d_z(x, y, z) \\
    0
\end{bmatrix} \tag{5.2} \]

5.4 Feature Extraction

As a result of deformable registration we obtain:

**Deformation Field** It contains information about differences in size, shape and locations of the corresponding anatomical structures of the brain images. In general, it corresponds to deformation that is needed in a voxel to map it to a reference image.

**Registered Image** This corresponds to registered images, after applying the deformation warp. Note that, even after deformable registration, there still
remains some residual differences between the reference image (Collin27) and the image that is being registered. We call the features extracted from the Deformation field as “Dynamic Features” and those extracted from the registered image as “Static Features”. This is because deformation field indicates how much a voxel has to be moved to make it similar to the reference image, while static features are extracted after the deformation warp has been applied.

A deformation field is a vector image that maps reference image voxel coordinates to the coordinates of the corresponding input image voxels. We extract 9 types of features from these vector fields: x, y, z components, theta, rho, phi angle components, length of the vectors(R), and the determinant of the Jacobian matrix of the deformation field. Combined with the registered intensity image we get 10 basic features. This way, for every deformation field and from the final registered image we obtain a number of 3D scalar images, one for each feature type. These images contain information about local differences in x, y, z coordinates and the distances between the corresponding voxels, as well as local contractions/expansions for every voxel neighborhood of the reference image. In order to capture this information with varying degree of locality, we create an image pyramid with 3 image scales for every scalar image obtained. First level in the image pyramid is the image itself, and every subsequent level is a smoothed and
subsampling (by the factor of two in each dimension) version of the previous level. At every level of the pyramid, we compute an asymmetry image. The asymmetry image is calculated by measuring the deviation from a perfectly symmetric brain. This is done in two steps. First a residual symmetry correction deformation field is calculated by registering the reference image to its completely symmetric counterpart. After a subject has been registered to the reference, we add this residual deformation to the registration deformation field. Finally the absolute value of the voxel-wise difference between voxel values on the left of MSP and their symmetric counterparts on the right are calculated. This is repeated for all feature types at all scales. Thus we have both local and asymmetry feature at 3 scales for each feature type. Figure 5.7 shows the local as well as the asymmetry feature type for each of these 10 features.

To capture local variations, we also compute neighborhood statistics for the voxels at various scales. We consider a 5x5x5 voxel neighborhood around each voxel and compute 19 different statistics. Figure 5.8 shows the statics images for local intensity.
Figure 5.7: 10 features extracted from the brain registration process and their asymmetry
Figure 5.8: 19 local statistical features
5.5 Results and Discussions

Human brain is one of the most complex structures in the universe. According to some estimates there are over 100 billion neurons in the human brain, and the number of neuro-circuits they form is astronomical. Thus understanding the working of the brain is a fairly challenging task. The problem is further complicated by the fact that the structure of the human brain is not static and varies across population. By applying our classification pipeline to the structural MR images, we try to find out regions of the brain that differentiate gender, age and some brain related diseases.

5.5.1 Gender Recognition

For this study we choose the Aging dataset as described earlier in section 5.1.1. The number of females in the dataset is 129 and the number of males is 132. The subjects were divided randomly into test and training sets, with same male-to-female ratio. 60% (128 subjects) of data were used for training was rest (128 subjects) were used for testing during each of the 100 random splits. AVR-mRMR-SFS feature selection combination was used with a LDA wrapper. 10000 features were retained after AVR ranking and 100 after mRMR ranking. The number of features selected after SFS varied in each iteration. Overall classification rate was optimized
Table 5.1: Gender classification results when different feature types were used. The last row shows the results when top 10000 features were chosen randomly instead of AVR ranking. Aging dataset was used for this experiment.

<table>
<thead>
<tr>
<th>Feature Type</th>
<th>Feature selection scheme</th>
<th>Sensitivity (Male)</th>
<th>Specificity (Female)</th>
<th>Overall</th>
</tr>
</thead>
<tbody>
<tr>
<td>Asymmetry</td>
<td>AVR+mRMR+SFS</td>
<td>81.2±6.4</td>
<td>88.0±4.8</td>
<td>84.5±5.1</td>
</tr>
<tr>
<td>Local</td>
<td>AVR+mRMR+SFS</td>
<td>84.5±4.3</td>
<td>89.5±4.2</td>
<td>86.5±4.3</td>
</tr>
<tr>
<td>All</td>
<td>AVR+mRMR+SFS</td>
<td>86.1±4.5</td>
<td>89.7±4.3</td>
<td>87.8±4.5</td>
</tr>
<tr>
<td>All</td>
<td>Random+mRMR+SFS</td>
<td>76.7±7.3</td>
<td>78.5±5.8</td>
<td>77.8±7.3</td>
</tr>
</tbody>
</table>

during the SFS stage. Results are shown in table 5.1. It can be seen that by using asymmetry features alone, above 80% classification can be achieved. Furthermore, adding Asymmetry feature to Local features boosted the classification rate by a significant 1.3% (p-Value < 0.001). The last row shows the classification results when top 10000 features to be input to mRMR stage were chosen randomly instead by AVR ranking. Table 5.10 to 5.27 show the statistics of feature types selected most during feature selection for gender classification.
Figure 5.9: Visualization of selected features for gender classification. The red blocks show the local features while the orange blocks show asymmetry features. The anatomical region where the features come from are also shown. Precuneus, Cerebelum and Pareital lobe were found to be important for gender classification.
Figure 5.10: Histogram of feature types selected after AVR ranking when all features were used for gender discrimination

Figure 5.11: Histogram of feature types selected after SFS when all features were used for gender discrimination
Figure 5.12: Histogram of feature types selected after AVR ranking when only local features were used for gender discrimination

Figure 5.13: Histogram of feature types selected after SFS when only local features were used for gender discrimination
Figure 5.14: Histogram of feature types selected after AVR ranking when only asymmetry features were used for gender discrimination

Figure 5.15: Histogram of feature types selected after SFS when only asymmetry features were used for gender discrimination
Figure 5.16: Histogram of important anatomical regions for gender classification when all the features were used

Figure 5.17: Histogram of anatomical regions selected after SFS when all features were used for gender discrimination
Figure 5.18: Histogram of important anatomical regions for gender classification when local the features were used

Figure 5.19: Histogram of important anatomical regions for after SFS when only local features were used for gender discrimination
Figure 5.20: Histogram of important anatomical regions for gender classification when only asymmetry features were used

Figure 5.21: Histogram of important anatomical regions for after SFS when only asymmetry features were used for gender discrimination
Figure 5.22: Histogram of feature stats selected after AVR when all features were used for gender discrimination

Figure 5.23: Histogram of feature stats selected after SFS when all features were used for gender discrimination
Figure 5.24: Histogram of feature stats selected after AVR when only local features were used for gender discrimination

Figure 5.25: Histogram of feature stats selected after SFS when only local features were used for gender discrimination
Figure 5.26: Histogram of feature stats selected after AVR when only asymmetry features were used for gender discrimination

Figure 5.27: Histogram of feature stats selected after SFS when only asymmetry features were used for gender discrimination
Figure 5.28: Visualization of important anatomical regions for gender classification when local features were used
Figure 5.29: Visualization of important anatomical regions for gender classification when local features were used
5.5.2 Age Estimation

Aging dataset was used of this experiment. Figure 5.1 shows the distribution of subjects in the aging dataset. For age estimation, we first selected top 10000 features based on AVR ranking by dividing the subjects into 5 years gaps and 6 classes. mRMR was then applied using continuous labels to select 100 features. This was followed by SFS with results of F-test as the selection criterion. Finally a linear regression model was trained on the selected features to minimize the mean square error. Note that the number of features finally selected varied between 40 to 60 at each iteration. There are 129 females and 132 males in this dataset. The training set consisted of 156 subjects while the test set consisted of 104 subjects, each of which were randomly chosen at each iteration. The process was repeated 100 times. We achieve a best mean squared error of 3.81±3.03 years when all the features were used. Figure 5.43 to 5.46 shows the ground truth and estimated age during 100 random split of the iteration when only local features, asymmetry features and all features were used. Table 5.2 summarizes the results for age classification.
<table>
<thead>
<tr>
<th>Feature Type</th>
<th>Absolute Training error (years)</th>
<th>Absolute Test error (years)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Asymmetry</td>
<td>2.8±2.3</td>
<td>4.8±3.4</td>
</tr>
<tr>
<td>Local</td>
<td>3.1±2.4</td>
<td>4.8±3.3</td>
</tr>
<tr>
<td>All</td>
<td>2.8±2.1</td>
<td>3.8±3.0</td>
</tr>
<tr>
<td>Random</td>
<td>5.2±3.5</td>
<td>5.7±3.9</td>
</tr>
</tbody>
</table>

Table 5.2: Age regression results when different feature types were used. The dataset was divided into 6 classes of 5 year range. AVR was used to select the top 10000 features. This was followed by mRMR using continuous labels to select top 100 features. SFS was then done with the selection criterion being the results of F-Test. The selected features were then used to train a linear regression model. The process was repeated 100 times with random split training/test set at each iteration. The last row shows the results when top 10000 features were chosen randomly instead of AVR ranking. Aging dataset was used for this experiment.
Figure 5.30: Visualization of selected features for age classification. The red blocks show the local features while the orange blocks show asymmetry features. The anatomical region where the features come from are also shown. The bottom row shows the features selected in a similar study by Teverovsky et al[119]. Teverovsky, while using only asymmetry feature found posterior horns of the lateral ventricles, the amygdala, the ventral putamen with a nearby region of the anterior inferior caudate nucleus, and the basal forebrain important for age estimation. In this study we found Caudate, Putamen, Insula, Thalmas and Hipocampus to be important for age estimation.
Figure 5.31: Histogram of feature types selected after SFS when only asymmetry features were used for age discrimination

Figure 5.32: Histogram of important anatomical regions for age classification after AVR ranking when all the features were used
Figure 5.33: Histogram of feature types selected after SFS when only asymmetry features were used for age discrimination

Figure 5.34: Histogram of important anatomical regions for age classification after AVR ranking when all the features were used
Figure 5.35: Histogram of feature types selected after SFS when only asymmetry features were used for age discrimination

Figure 5.36: Histogram of important anatomical regions for age classification after Rand ranking when all the features were used
Figure 5.37: Histogram of feature types selected after SFS when only asymmetry features were used for age discrimination.

Figure 5.38: Histogram of important anatomical regions for age classification after Rand ranking when all the features were used.
Figure 5.39: Histogram of feature types selected after SFS when only asymmetry features were used for age discrimination

Figure 5.40: Histogram of important anatomical regions for age classification after AVR ranking when Asymmetry the features were used
Figure 5.41: Histogram of feature types selected after SFS when only asymmetry features were used for age discrimination

Figure 5.42: Histogram of important anatomical regions for age classification after AVR ranking when Asymmetry the features were used
Figure 5.43: Histogram of feature types selected after SFS when only asymmetry features were used for age discrimination.

Figure 5.44: Histogram of important anatomical regions for age classification after AVR ranking when Local the features were used.
Figure 5.45: Histogram of feature types selected after SFS when only asymmetry features were used for age discrimination

Figure 5.46: Histogram of important anatomical regions for age classification after AVR ranking when Local the features were used
Figure 5.47: Histogram of feature types selected after AVR ranking when all features were used for age discrimination

Figure 5.48: Histogram of feature types selected after SFS when all features were used for age discrimination
Figure 5.49: Histogram of feature types selected after AVR ranking when only local features were used for age discrimination

Figure 5.50: Histogram of feature types selected after SFS when only local features were used for age discrimination
Figure 5.51: Histogram of feature types selected after AVR ranking when only asymmetry features were used for age discrimination

Figure 5.52: Histogram of feature types selected after SFS when only asymmetry features were used for age discrimination
Figure 5.53: Histogram of important anatomical regions for age classification after AVR ranking when all the features were used

Figure 5.54: Histogram of anatomical regions selected after SFS when all features were used for age discrimination
Figure 5.55: Histogram of important anatomical regions for age classification after AVR ranking when local the features were used

Figure 5.56: Histogram of important anatomical regions for after SFS when only local features were used for age discrimination
Figure 5.57: Histogram of important anatomical regions for age classification after AVR ranking when only asymmetry features were used

Figure 5.58: Histogram of important anatomical regions for after SFS when only asymmetry features were used for age discrimination
Figure 5.59: Histogram of feature stats selected after AVR when all features were used for age discrimination

Figure 5.60: Histogram of feature stats selected after SFS when all features were used for age discrimination
Figure 5.61: Histogram of feature stats selected after AVR when only local features were used for age discrimination

Figure 5.62: Histogram of feature stats selected after SFS when only local features were used for age discrimination
Figure 5.63: Histogram of feature stats selected after AVR when only asymmetry features were used for age discrimination

Figure 5.64: Histogram of feature stats selected after SFS when only asymmetry features were used for age discrimination
Table 5.3: Results of leave 10 out Classification on ADRC dataset averaged over 100 iteration. The row indicates the true class while column indicates the classified class. We used AVR+mRMR+SFS followed by LDA classifier to get the above results.

5.5.3 Computer Aided Diagnosis (CAD) for Alzheimer’s Disease

5.5.3.1 ADRC

The ADRC dataset consists of 3 classes. Control, Mild Cognitive Impaired (MCI) and Alzheimer’s Diagnosed (AD) with 20, 26 and 20 subjects in respective categories. We do leave-10-out cross validation with 100 random splits. The classification results are shown in Table 5.3. Figure 5.65 shows the distribution of subjects in top 3 selected feature dimensions during on random split. Figure 5.67 to 5.84 show the statistics of features selected after AVR and SFS stage for 100 iterations.
Figure 5.65: Feature Distribution in top 3 selected feature space during one random split for ADRC dataset
Figure 5.66: Visualization of selected features for Alzheimer’s classification. The red blocks show the local features while the orange blocks show asymmetry features. The anatomical region where the features come from are also shown. The bottom row shows the features selected in a similar study by Liu et al.[83]. While in this study we treat the problem as 3 class problem, Liu et al. treated this problem as 3 separate binary class problem (AD/MCI, MCI/CTL, CTL/AD). Liu et al. found Isthmus region, Left ventral striatum, right Uncinate Fasciculus, posterior temporal lobe and the cingulate important for Alzheimer’s disease classification. In this study we found Parietal lobe, occipital lobe and mid temporal lobe important for Alzheimer’s disease classification.
Figure 5.67: Histogram of all feature types for alzheimer’s disease discrimination after AVR ranking on the ADRC dataset after 100 iterations when all features were used.

Figure 5.68: Histogram of all feature types for alzheimer’s disease discrimination after SFS on the ADRC dataset after 100 iterations when all features were used.
Figure 5.69: Histogram of all feature types for alzheimer’s disease discrimination after AVR ranking on the ADRC dataset after 100 iterations when only local features were used.
Figure 5.70: Histogram of all feature types for alzheimer’s disease discrimination after SFS on the ADRC dataset after 100 iterations when only local features were used.
Figure 5.71: Histogram of all feature types for alzheimer’s disease discrimination after AVR ranking on the ADRC dataset after 100 iterations when only asymmetry features were used.
Figure 5.72: Histogram of all feature types for alzheimer’s disease discrimination after SFS on the ADRC dataset after 100 iterations when only asymmetry features were used.
Figure 5.73: Histogram of top 21 anatomical regions selected for Alzheimer’s disease discrimination after AVR ranking on the ADRC dataset after 100 iterations when all features were used.
Figure 5.74: Histogram of top 21 anatomical regions selected for alzheimer’s disease discrimination after SFS on the ADRC dataset after 100 iterations when all features were used.
Figure 5.75: Histogram of top 21 anatomical regions selected for alzheimer’s disease discrimination after AVR ranking on the ADRC dataset after 100 iterations when only local features were used.
Figure 5.76: Histogram of top 21 anatomical regions selected for Alzheimer's disease discrimination after SFS on the ADRC dataset after 100 iterations when only local features were used.
Figure 5.77: Histogram of top 21 anatomical regions selected for alzheimer’s disease discrimination after AVR ranking on the ADRC dataset after 100 iterations when only asymmetry features were used.
Figure 5.78: Histogram of top 21 anatomical regions selected for Alzheimer’s disease discrimination after SFS on the ADRC dataset after 100 iterations when only asymmetry features were used.
Figure 5.79: Histogram of all feature stats for alzheimer’s disease discrimination after AVR ranking on the ADRC dataset after 100 iterations when all features were used.

Figure 5.80: Histogram of all feature stats for alzheimer’s disease discrimination after SFS on the ADRC dataset after 100 iterations when all features were used.
Figure 5.81: Histogram of all feature stats for Alzheimer's disease discrimination after AVR ranking on the ADRC dataset after 100 iterations when only local features were used.
Figure 5.82: Histogram of all feature stats for alzheimer’s disease discrimination after SFS on the ADRC dataset after 100 iterations when only local features were used.
Figure 5.83: Histogram of all feature stats for alzheimer’s disease discrimination after AVR ranking on the ADRC dataset after 100 iterations when only asymmetry features were used.
Figure 5.84: Histogram of all feature stats for Alzheimer’s disease discrimination after SFS on the ADRC dataset after 100 iterations when only asymmetry features were used.
<table>
<thead>
<tr>
<th></th>
<th>Control</th>
<th>MCI</th>
<th>AD</th>
</tr>
</thead>
<tbody>
<tr>
<td>CHS 1</td>
<td>125</td>
<td>44</td>
<td>35</td>
</tr>
<tr>
<td>CHS 2</td>
<td>34</td>
<td>19</td>
<td>18</td>
</tr>
<tr>
<td>CHS 3</td>
<td>230</td>
<td>80</td>
<td>21</td>
</tr>
</tbody>
</table>

Table 5.4: Table shows the subject distribution in CHS datasets for whom diagnosis is available.

<table>
<thead>
<tr>
<th></th>
<th>Control</th>
<th>MCI</th>
<th>AD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Control</td>
<td>60.00 ± 18.7</td>
<td>25.94 ± 16.4</td>
<td>14.05 ± 17.5</td>
</tr>
<tr>
<td>MCI</td>
<td>21.00 ± 31.6</td>
<td>57.00 ± 25.6</td>
<td>22.00 ± 27.1</td>
</tr>
<tr>
<td>AD</td>
<td>13.84 ± 22.1</td>
<td>19.99 ± 18.7</td>
<td>66.15 ± 21.0</td>
</tr>
</tbody>
</table>

Table 5.5: Results of leave 10 out Classification on CHS 1 dataset averaged over 100 iteration. The row indicates the true class while column indicates the classified class. We used AVR+mRMR+SFS followed by LDA classifier to get the above results.

5.5.3.2 CHS

Table 5.4 shows the number of subjects in each category in the CHS datasets. The results of classification on CHS datasets are tabulated in table 5.9. A high degree of over training was observed in the CHS datasets. No significant difference in classification rate was observed among the CHS datasets. Figure 5.85 to 5.102 show the statistics of features selected in the CHS datasets.

<table>
<thead>
<tr>
<th></th>
<th>Control</th>
<th>MCI</th>
<th>AD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Control</td>
<td>63.33 ± 32.4</td>
<td>25.00 ± 19.6</td>
<td>11.46 ± 25.3</td>
</tr>
<tr>
<td>MCI</td>
<td>19.99 ± 23.3</td>
<td>56.66 ± 39.1</td>
<td>23.33 ± 31.6</td>
</tr>
<tr>
<td>AD</td>
<td>06.66 ± 21.0</td>
<td>36.66 ± 18.9</td>
<td>56.66 ± 44.5</td>
</tr>
</tbody>
</table>

Table 5.6: Results of leave 10 out Classification on CHS 2 dataset averaged over 100 iteration. The row indicates the true class while column indicates the classified class. We used AVR+mRMR+SFS followed by LDA classifier to get the above results.
Table 5.7: Results of leave 10 out Classification on CHS 3 dataset averaged over 100 iteration. The row indicates the true class while column indicates the classified class. We used AVR+mRMR+SFS followed by LDA classifier to get the above results.

<table>
<thead>
<tr>
<th></th>
<th>Control</th>
<th>MCI</th>
<th>AD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Control</td>
<td>73.47 ± 10.0</td>
<td>21.73 ± 38.3</td>
<td>04.28 ± 38.3</td>
</tr>
<tr>
<td>MCI</td>
<td>37.50 ± 18.8</td>
<td>56.25 ± 15.8</td>
<td>06.25 ± 20.6</td>
</tr>
<tr>
<td>AD</td>
<td>01.00 ± 14.7</td>
<td>44.00 ± 19.1</td>
<td>54.10 ± 20.1</td>
</tr>
</tbody>
</table>

Table 5.8: Results of leave 10 out Classification on CHS dataset averaged over 100 iteration. LDA was used as the final classifier. While the number of subjects used for training varied for different classes, 3 subjects from each class and 1 randomly chosen subject were used for testing at each iteration. A high standard deviations indicates the instability of the results.

<table>
<thead>
<tr>
<th></th>
<th>CHS 1</th>
<th>CHS 2</th>
<th>CHS 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Asymmetry Train</td>
<td>100.0±0.0</td>
<td>100.0±0.0</td>
<td>100.0±0.0</td>
</tr>
<tr>
<td>Asymmetry Test</td>
<td>54.0±42.3</td>
<td>39.8±45.2</td>
<td>49.9±46.3</td>
</tr>
<tr>
<td>Local Train</td>
<td>100.0±0.0</td>
<td>100.0±0.0</td>
<td>100.0±0.0</td>
</tr>
<tr>
<td>Local Test</td>
<td>59.9±30.5</td>
<td>51.3±34.0</td>
<td>53.1±39.0</td>
</tr>
<tr>
<td>All Train</td>
<td>100.0±0.0</td>
<td>100.0±0.0</td>
<td>100.0±0.0</td>
</tr>
<tr>
<td>All Test</td>
<td>62.0±24.5</td>
<td>59.0±40.0</td>
<td>67.0±45.3</td>
</tr>
</tbody>
</table>

Table 5.9: Results of Classification on CHS dataset averaged over 100 iteration when equal number of subjects were used in training and test. A high standard deviations indicates the instability of the results.

<table>
<thead>
<tr>
<th></th>
<th>CHS 1</th>
<th>CHS 2</th>
<th>CHS 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Asymmetry Train</td>
<td>100.0±0.0</td>
<td>100.0±0.0</td>
<td>100.0±0.0</td>
</tr>
<tr>
<td>Asymmetry Test</td>
<td>30.0±60.3</td>
<td>51.5±30.0</td>
<td>49.9±50.0</td>
</tr>
<tr>
<td>Local Train</td>
<td>100.0±0.0</td>
<td>100.0±0.0</td>
<td>100.0±0.0</td>
</tr>
<tr>
<td>Local Test</td>
<td>35.7±40.5</td>
<td>40.0±52.0</td>
<td>43.5±37.0</td>
</tr>
<tr>
<td>All Train</td>
<td>100.0±0.0</td>
<td>100.0±0.0</td>
<td>100.0±0.0</td>
</tr>
<tr>
<td>All Test</td>
<td>38.0±51.5</td>
<td>41.0±50.0</td>
<td>60.0±35.0</td>
</tr>
</tbody>
</table>
Figure 5.85: Histogram of all feature types for CHS1 disease discrimination after AVR ranking on the CHS1 dataset after 100 iterations when all features were used.

Figure 5.86: Histogram of all feature types for CHS1 disease discrimination after SFS on the CHS1 dataset after 100 iterations when all features were used.
Figure 5.87: Histogram of top 21 anatomical regions selected for CHS1 disease discrimination after AVR ranking on the CHS1 dataset after 100 iterations when all features were used.
Figure 5.88: Histogram of top 21 anatomical regions selected for CHS1 disease discrimination after SFS on the CHS1 dataset after 100 iterations when all features were used.
Figure 5.89: Histogram of all feature stats for CHS1 disease discrimination after AVR ranking on the CHS1 dataset after 100 iterations when all features were used.

Figure 5.90: Histogram of all feature stats for CHS1 disease discrimination after SFS on the CHS1 dataset after 100 iterations when all features were used.
Figure 5.91: Histogram of all feature types for CHS2 disease discrimination after AVR ranking on the CHS2 dataset after 100 iterations when all features were used.

Figure 5.92: Histogram of all feature types for CHS2 disease discrimination after SFS on the CHS2 dataset after 100 iterations when all features were used.
Figure 5.93: Histogram of top 21 anatomical regions selected for CHS2 disease discrimination after AVR ranking on the CHS2 dataset after 100 iterations when all features were used.
Figure 5.94: Histogram of top 21 anatomical regions selected for CHS2 disease discrimination after SFS on the CHS2 dataset after 100 iterations when all features were used.
Figure 5.95: Histogram of all feature stats for CHS2 disease discrimination after AVR ranking on the CHS2 dataset after 100 iterations when all features were used.

Figure 5.96: Histogram of all feature stats for CHS2 disease discrimination after SFS on the CHS2 dataset after 100 iterations when all features were used.
Figure 5.97: Histogram of all feature types for CHS3 disease discrimination after AVR ranking on the CHS3 dataset after 100 iterations when all features were used.

Figure 5.98: Histogram of all feature types for CHS3 disease discrimination after SFS on the CHS3 dataset after 100 iterations when all features were used.
Figure 5.99: Histogram of top 21 anatomical regions selected for CHS3 disease discrimination after AVR ranking on the CHS3 dataset after 100 iterations when all features were used.

Figure 5.100: Histogram of top 21 anatomical regions selected for CHS3 disease discrimination after SFS on the CHS3 dataset after 100 iterations when all features were used.
Figure 5.101: Histogram of all feature stats for CHS3 disease discrimination after AVR ranking on the CHS3 dataset after 100 iterations when all features were used.

Figure 5.102: Histogram of all feature stats for CHS3 disease discrimination after SFS on the CHS3 dataset after 100 iterations when all features were used.
Chapter 6

Summary and Conclusion

In this work, we proposed, implemented and evaluated a classification framework with huge feature dimensions. We also proposed some new novel features for plant, 3D face and volumetric MRI data. The effectiveness of the pipeline was demonstrated by various discriminative experiments on these datasets, including gender, expression, disease, plant species classification. A summary of the work done in this work is given below.

Classification pipeline We successfully showed the capability of the pipeline for handling features as high as 90 million. Also by achieving competitive results on a variety of bio-medical datasets, we showed the generic capabilities of the pipeline. Apart from the classification rate, the pipeline also outputs the importance of individual features for a
discrimination task. By mapping these back spatially we get a lot more insight than by just having a classification rate. Figure 6.1 shows the output of the pipeline for a particular experiment. The user can specify various parameters before the experiment and the pipeline automatically prepares a summary report with various figures and graphs. The pipeline can also be run in batch mode and the pipeline would prepare the report of the experiments. Thus the pipeline saves user a lot of time and effort. The modularity of the pipeline helps user to methodically experiment with a lot of parameters and study their effects.

**Plant leaf analysis** We proposed features based on leaf shape and texture (more than 3,000 features) for plant leaf analysis and used those features for automatic plant species and Ecotypes classification and also identified corresponding important features. By identifying the genes responsible for these features, physiologists would get better understanding of the plant growth model.

**3D Face discrimination** We proposed features based on surface height, orientation and curvature (above 300,000) for 3D facial analysis and applied it to the problems of gender, age, expression and person identification. We did an in-depth study of inter/intra race gender differences and identify regions of the face which differ between the groups.
We have achieved the best classification rate reported so far in literature for 3D facial gender classification and age prediction from 3D face. We also achieved competitive classification rates for expression and person identification from 3D mesh. We also identify which features are important for various face discrimination tasks. Identifying regions of the face which discriminate gender, age, race etc is of particular interest to many psychologists and anthropologists.

3D Brain analysis We build on the work done in [119, 120, 83] to extract many novel features from the brain MR images. We use features extracted from the deformation field and the intensity (above 90 million) to many novel applications such as gender and age estimation from brain. We further identified which regions of the brain are more important for a particular discrimination task. This provides a better understanding of the human brain. We also applied our framework for identifying regions of the brain which are affected by Alzheimer’s disease and developed a framework which could be used for computer-aided-diagnosis of neuro-generative diseases.

Importance of Asymmetry Following up on the work of [81, 79, 80, 88, 117, 120, 83, 119] on quantified face and brain asymmetry and using an improved Mid-Sagittal Plane (MSP) extraction algorithm [121], we investigated the importance of asymmetry features for 3D face and
Figure 6.1: The figure shows the screenshot of the output of the pipeline for an experiment. The user can specify various parameters before the experiment and the pipeline automatically prepares a summary report with various figures and graphs. The pipeline can also be run in batch mode and the pipeline would prepares the report of the experiments. Thus the pipeline saves user a lot of time and effort. The modularity of the pipeline helps user to methodically experiments with a lot of parameters and study their effects.

volumetric brain datasets. We successfully showed that symmetry is a very important feature for discrimination tasks involving face and brain. We also showed that by combining asymmetry features with local features we achieve better classification results than by using just local features alone.

Table 6.1 summarizes the results of various experiments.
<table>
<thead>
<tr>
<th>Domain</th>
<th>Image Type</th>
<th>IFS</th>
<th>FFS</th>
<th>Discriminant</th>
<th>Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>Plant specie</td>
<td>2D color image</td>
<td>10^4</td>
<td>20-30</td>
<td>LDA</td>
<td>98.93(±2.3)%</td>
</tr>
<tr>
<td>Face Gender</td>
<td>3D surface mesh</td>
<td>10^6</td>
<td>40-60</td>
<td>LDA</td>
<td>96.3(±3.5) %</td>
</tr>
<tr>
<td>Age Classification from face</td>
<td>3D surface mesh</td>
<td>10^6</td>
<td>20-30</td>
<td>LDA</td>
<td>88.8(±5.7) %</td>
</tr>
<tr>
<td></td>
<td>3D surface mesh</td>
<td>10^6</td>
<td>40-60</td>
<td>Linear</td>
<td>4.0(±3.4) years</td>
</tr>
<tr>
<td>Face Expression</td>
<td>3D surface mesh</td>
<td>10^6</td>
<td>50-90</td>
<td>LDA</td>
<td>88.87(±3.1) %</td>
</tr>
<tr>
<td>Brain Gender</td>
<td>Structural volumetric MRI</td>
<td>10^8</td>
<td>40-60</td>
<td>LDA</td>
<td>87.8(±4.5) %</td>
</tr>
<tr>
<td>Brain Age</td>
<td>Structural volumetric MRI</td>
<td>10^8</td>
<td>40-60</td>
<td>Linear</td>
<td>3.8(±3.0) years</td>
</tr>
<tr>
<td>ADRC</td>
<td>Structural volumetric MRI</td>
<td>10^8</td>
<td>10-25</td>
<td>LDA</td>
<td>87.4(±15.0) %</td>
</tr>
<tr>
<td>CHS (average)</td>
<td>Structural volumetric MRI</td>
<td>10^8</td>
<td>5-20</td>
<td>LDA</td>
<td>64.6(±40.5) %</td>
</tr>
</tbody>
</table>

Table 6.1: Input image type, initial feature set size, average selected feature set size and selected classification results. For all the experiments, AVR was used for univariate ranking, followed by mRMR and SFS feature selection with a LDA wrapper. Note IFS= Initial feature set size, FFS= Final feature set size
Bibliography


[55] D. Kim and H. Oh. Cvtresh: R package for level-dependent cross-
2006.


[58] R. Kohavi. *A Study of Cross-Validation and Bootstrap for Accuracy Es-


Kaufmann, 1996.

[61] Y. Kondou, M. Higuchi, and M. Matsui. High-throughput characterization
of plant gene functions by using gain-of-function technology. *Annual

[62] D. Kontos, V. Megalooikonomou, and J. C. Gee. Morphometric analy-
thesis of brain images with reduced number of statistical tests: A study
on the gender-related differentiation of the corpus callosum. *Artificial

[63] D. Le and S. Satoh. Robust object detection using fast feature selection
from huge feature sets. In *Image Processing, 2006 IEEE International


[91] I. Mpiperis, S. Malassiotis, V. Petridis, and M.G. Strintzis. 3d facial expression recognition using swarm intelligence. *Acoustics, Speech and


[143] Y. Zhang, C. Ding, and T. Li. Gene selection algorithm by combining reliefF and mRMR. BMC GENOMICS, 9(Suppl. 2), 2007.

Appendix A

Moments

Moments can be used to compactly describe the nature of distribution of a function such as its mean, skewness etc. Let $z$ be a random variable denoting the values taken by a discrete function and $p(z_i)$, $i=0,...,N$, be the corresponding histogram, where $N$ is number of histogram bins. Then, the $n^{th}$ moment of $z$ about its mean is:

$$\mu_n(z) = \sum_{i=0}^{N-1} (z_i - m)^n p(z_i) \quad (A.1)$$

where $m$ is the mean of $z$, given by

$$m = \sum_{i=0}^{N-1} z_ip(z_i) \quad (A.2)$$

The second moment is the variance and its square root is the standard
deviation $\sigma$. The normalized $n^{th}$ moment of $z$ about its mean is given by

$$\mu_n(z) = \frac{\sum_{i=0}^{N-1} (z_i - m)^n p(z_i)}{\sigma^n} \quad (A.3)$$

The normalized third central moment is called skewness and is a measure of lopsidedness of the histogram. Similarly fourth normalized moment (also called Kurtosis) is a measure of relative flatness. Many other useful features can be derived from moments. Some of them are listed below.

$$Smoothness = 1 - \frac{1}{1 + \sigma^2(z)} \quad (A.4)$$

$$Uniformity = \sum_{i=0}^{N-1} p^2(z_i) \quad (A.5)$$

$$Entropy = -\sum_{i=0}^{N-1} p(z_i) \log_2 p(z_i) \quad (A.6)$$

For a two dimensional function (e.g. contour of a leaf), we can further extract moments invariants which are insensitive to scale, rotation or translation. For a 2D-function $f(x,y)$, the central moment of order $(p+q)$ is defined as
\[ \mu_{pq} = \sum_x \sum_y (x - \bar{x})^p (y - \bar{y})^q \]  \hspace{1cm} (A.7)

The normalized central moments of order \((p+q)\) is defined as

\[ \eta_{pq} = \frac{\mu_{pq}}{\mu_{00}} \]  \hspace{1cm} (A.8)

where

\[ \gamma = \frac{p + q}{2} + 1 \]  \hspace{1cm} (A.9)

It can be shown that the following set of invariant moments can be derived from second and third moments. These are invariant to translation, rotation and scale changes [47].

\[ \phi_1 = \eta_{20} + \eta_{02} \]

\[ \phi_2 = (\eta_{20} - \eta_{02})^2 + 4\eta_{11}^2 \]

\[ \phi_3 = (\eta_{30} - 3\eta_{12})^2 + (3\eta_{21} - \eta_{03})^2 \]

\[ \phi_4 = (\eta_{30} + \eta_{12})^2 + (\eta_{21} + \eta_{03})^2 \]
\[
\phi_5 = (\eta_{30} - 3\eta_{12})(\eta_{30} + \eta_{12})[(\eta_{30} + \eta_{12})^2 - 3(\eta_{21} - \eta_{03})^2] + \\
(3\eta_{21} - \eta_{03})(\eta_{21} + \eta_{03})[3(\eta_{30} + \eta_{12})^2 - (\eta_{21} + \eta_{03})^2]
\]

\[
\phi_6 = (\eta_{20} - \eta_{02})[(\eta_{30} + \eta_{12})^2 - (\eta_{21} + \eta_{03})^2] + 4\eta_{11}(\eta_{30} + \eta_{12})(\eta_{21} + \eta_{03})
\]

\[
\phi_7 = (3\eta_{21} - \eta_{03})(\eta_{30} + \eta_{12})[(\eta_{30} + \eta_{12})^2 - 3(\eta_{21} - \eta_{03})^2] + \\
(3\eta_{21} - \eta_{03})(\eta_{21} + \eta_{03})[3(\eta_{30} + \eta_{12})^2 - (\eta_{21} + \eta_{03})^2]
\]

The concept can be generalized to higher dimensions too. More details can be found in [20, 36, 47].
Appendix B

Supplementary Materials

The supplementary material of the thesis can be accessed at the following URL: http://somesh.net/msthesis/