Contents

1 Introduction **5**
1.1 Organization of this Thesis 7

2 Exploratory Data Analysis **9**
2.1 Knowledge Discovery 10
2.2 Visual Data Mining 11
2.2.1 Scientific Visualization 13
2.2.2 The Process of Scientific Visualization 14
2.2.3 Classification Schemes for Scientific Data 15
2.2.4 Scientific Visualization Techniques 17
2.2.4.1 Two-Dimensional Domain 18
2.2.4.2 Three-Dimensional Domain 21
2.2.5 Color in Visualization 24
2.2.5.1 Color Fundamentals 25
2.2.5.2 Device Independent Color Spaces 26
2.2.5.3 Device Dependent Color Spaces 27
2.2.5.4 Color Scale Properties 28
2.2.5.5 Univariate Color Scales 29
2.2.5.6 Perceptually Optimized Color Mapping Techniques . . . 30
2.2.5.7 CIELUV Color Scales Based on sRGB Color Specifications 32
2.2.5.8 Bivariate and Trivariate Color Scales 33
2.3 Image Fusion 33
2.3.1 Dimensionality Reduction 36
2.3.2 Feature Selection 38
2.3.3 Feature Extraction 38
2.4 Conclusion 40

3 Linear Transformations **41**
3.1 Principal Component Analysis 42
3.1.1 Intrinsic Data Dimensionality 45
3.1.2 Eigenvector Computation Using Neural Networks 46
3.1.3 Principal Component Analysis and Image Fusion 47
3.1.4 Case Study: DCE-MRI of the Breast 49
3.1.4.1 Motivation 49
3.1.4.2 Detection and Classification of Lesions 51
3.1.4.3 Computer Aided Diagnosis 53

3.1.4.4 Material and Methods . . . 55
3.1.4.5 Preparation of the Data . . . 56
3.1.4.6 Experimental Results . . . 57
3.2 Procrustes Analysis . . . 70
3.2.1 Comparison of Point Configurations . . . 70
3.2.2 Comparison of Subspaces . . . 71
3.2.3 Case Study Revisited: DCE-MRI of the Breast . . . 73
3.2.3.1 Experimental Results . . . 73
3.3 Independent Component Analysis . . . 77
3.3.1 Statistical Independence . . . 78
3.3.2 Non-Gaussianity . . . 79
3.3.3 The FastICA Algorithm . . . 81
3.3.4 Validation of Independent Components . . . 82
3.3.5 ICA and Image Fusion . . . 86
3.3.6 Case Study Revisited: DCE-MRI of the Breast . . . 88
3.3.6.1 Experimental Results . . . 89
3.4 Conclusion . . . 97

4 Non-Linear Transformations 99
4.1 Self-Organizing Maps . . . 100
4.1.1 The Incremental Algorithm . . . 101
4.1.2 Topology Preservation . . . 103
4.1.3 Intrinsic Data Dimensionality . . . 106
4.1.4 SOM Based Data Visualization . . . 108
4.1.4.1 Analysis of the Structure of a Data Distribution . . . 108
4.1.4.2 Reference Vector Based Analysis . . . 109
4.1.4.3 SOM Calibration . . . 109
4.1.5 Image Fusion and Self-Organizing Maps . . . 110
4.1.6 Case Study: Screening of Synthetic Stationary-Phase Promoters . 111
4.1.6.1 Motivation . . . 111
4.1.6.2 Material and Method . . . 112
4.1.6.3 Experimental Results . . . 115
4.2 Hyperbolic Self-Organizing Maps . . . 118
4.2.1 Characteristics of Hyperbolic Spaces . . . 119
4.2.2 The Poincaré Disc Model . . . 121
4.2.3 A Hyperbolic Lattice and a Modified Adaptation Method . . . 121
4.2.4 Visualizing Hyperbolic Self-Organizing Maps . . . 122
4.2.5 The Hyperbolic Self-Organizing Maps in Image Fusion . . . 123
4.2.6 Case Study Revisited: DCE-MRI of the Breast . . . 124
4.2.6.1 Experimental Results . . . 124
4.3 Topographic Mapping for Proximity Data . . . 135
4.3.1 Probabilistic Autoencoders . . . 136
4.3.2 EM Algorithm and Deterministic Annealing for Optimization . . . 137

4.3.2.1 An Approximation of the Optimization Scheme 138
4.3.2.2 A Hyperbolic Topographic Mapping for Proximity Data . 138
4.3.3 Case Study Revisited: DCE-MRI of the Breast 140
4.3.3.1 Material and Methods 141
4.3.3.2 Experimental Results for the Exploration of Randomly Sampled Image Sections 143
4.3.3.3 Experimental Results for Image Fusion based on Proximity Data 147
4.4 Conclusion 155

5 Summary and Outlook **157**

6 Appendix **161**
6.1 Performance Metrics 161

Bibliography **167**