## Contents

Preface xvii  
List of Contributors xix  

### CHAPTER 1  Humanistic Intelligence: “Wear Comp” As a New Framework and Application for Intelligent Signal Processing  

1  

**INTRODUCTION**  
- Why Humanistic Intelligence?  2  
- Humanistic Intelligence Does Not Necessarily Mean “user-friendly”  2  

**“WEAR COMP” AS MEANS OF REALIZING HUMANISTIC INTELLIGENCE**  3  
- Basic Principals of WearComp  3  
  - **Operational modes of WearComp**  3  
- The Six Basic Signal Flow Paths of WearComp  5  

**PHILOSOPHICAL ISSUES**  6  
- Fundamental Issues of WearComp  6  
  - **Historical context**  6  
  - The shift from guns to cameras and computers  6  
  - The shift from draconian punishment to micro management  7  
  - **Fundamental issues of WearComp**  8  
- Aspects of WearComp and Personal Empowerment  9  

**PRACTICAL EMBODIMENT OF WEARCOMP**  10  
- Building Signal-Processing Devices Directly into Fabric  12  
  - Remaining issue of underwearable signal processing hardware  13  
- Multidimensional Signal Input for Humanistic Intelligence  14  
  - **Safety first**  15  
  - More than just a health status monitor  16  

**PERSONAL IMAGING APPLICATION OF HUMANISTIC INTELLIGENCE**  16  

- Some Simple Illustrative Examples  16  
  - Always ready: From point and click to “look and think”  16  
  - Personal safety device for reducing crime  17  
  - The retro-autofocus example: Human in the signal processing loop  17  
- Mathematical Framework for Personal Imaging  18  
  - **Quantigraphic imaging and the Wyckoff Principle**  18  
  - **Video orbits**  20  

vii
Contents

Dynamic range and “dynamic domain” 24
Bi-foveated WearCam 30
Lightspace modeling for HI 31
BEYOND VIDEO: SYNTHETIC SYNESTHESIA AND PERSONAL IMAGING 34
Synthetic Synesthesia: Adding New Sensory Capabilities to the Body 34
Safety first 35
A true extension of the mind and body 36
CONCLUSIONS 36
ACKNOWLEDGMENTS 37
REFERENCES 37

CHAPTER 2  Adaptive Stochastic Resonance 40

ABSTRACT 40
STOCHASTIC RESONANCE AND ADAPTIVE FUNCTION APPROXIMATION 41
SR DYNAMICAL SYSTEMS 46
SR PERFORMANCE MEASURES 51
ADDITIVE FUZZY SYSTEMS AND FUNCTION APPROXIMATION 56
SR LEARNING AND EQUILIBRIUM 58
The Signal-to-Noise Ratio in Nonlinear Systems 58
Supervised Gradient Learning and SR Optimality 63
SR LEARNING: SIMULATION RESULTS 68
SR Test Case: The Quartic Bistable System 70
Other SR Test Cases 77
Fuzzy Sr Learning: The Quartic Bistable System 81
CONCLUSIONS 86
REFERENCES 89
APPENDIX A. THE STANDARD ADDITIVE MODEL (SAM) THEOREM 104
APPENDIX B. SAM GRADIENT LEARNING 106

CHAPTER 3  Learning in the Presence of Noise 108

ABSTRACT 108
INTRODUCTION 108
FINANCIAL TIME SERIES PREDICTION 110
IMPACT OF NOISE ON LEARNING 111
The Learning Problem 111
Performance of a Learning System 113
Estimating the Model Limitation 116
APPLICATION TO FINANCIAL MARKET FORECASTING 116
CONCLUSION 117
ACKNOWLEDGMENTS 118
APPENDIX 119
REFERENCES 125

CHAPTER 4  Incorporating Prior Information in Machine Learning by Creating Virtual Examples 127

ABSTRACT 127
LEARNING FROM EXAMPLES 128
Background: Learning as Function Approximation 128
PRIOR INFORMATION AND THE PROBLEM OF SAMPLE COMPLEXITY 129
VIRTUAL EXAMPLES: A FRAMEWORK FOR PRIOR INFORMATION 133
The General Framework 134
Techniques for Prior Information and Related Research 135
  Prior Knowledge in the Choice of Variables of Features 135
  Prior Knowledge in the Learning Technique 135
  Generating New Examples with Prior Knowledge 136
  Incorporating Prior Knowledge as Hints 137
VIRTUAL EXAMPLES AND REGULARIZATION 138
  Regularization Theory and RBF 138
  Regularization Theory in Presence of Radial Symmetry 140
  Radial Symmetry and “Virtual” Examples 141
VIRTUAL EXAMPLES IN VISION AND SPEECH 143
  Virtual Views for Object Recognition 144
  Symmetry as Prior Information 146
  More General Transformations: Linear Object Classes 147
    3D Objects, 2D Projections, and Linear Classes 148
    Implications 149
    Learning the Transformation 150
  Virtual Examples in Speech Recognition 152
CONCLUSIONS 158
REFERENCES 158

CHAPTER 5 Deterministic Annealing for Clustering, Compression, Classification, Regression, and Speech Recognition 163
ABSTRACT 163
INTRODUCTION 163
DETERMINISTIC ANNEALING FOR UNSUPERVISED LEARNING 166
  Clustering 166
    Principled Derivation of Deterministic Annealing 167
    Statistical Physics Analogy 169
    Mass-Constrained Clustering 172
    Preferred Implementation of the DA Clustering Algorithm 177
    Illustrative Examples 178
  Extensions and Applications 180
    Vector Quantization for Noisy Channels 180
    Entropy-Constrained Vector Quantizer Design 182
    Structurally Constrained Vector Quantizer Design 183
    Graph-Theoretic and Other Optimization Problems 185
DETERMINISTIC ANNEALING FOR SUPERVISED LEARNING 188
  Problem Formulation 188
  Basic Derivation 190
  Generality and Wide Applicability of the DA Solution 192
    Regression, Classification, and Clustering 192
    Structures 193
  Experimental Results 197
    VQ Classifier Design 197
    RBF Classifier Design 201
    MLP Classifier Design 201
    Piecewise Regression 202
    Mixture of Experts 206
Contents

CHAPTER 8

Semiparametric Support Vector Machines for Nonlinear Model Estimation

ABSTRACT 295
INTRODUCTION 295
PROBLEM SETTING 296
OPTIMUM LINE SEARCH FOR A GENERIC DIRECTION 297
SIMPLE ALGORITHMS FOR PURELY NONPARAMETRIC APPROACH 298
A Gradient-Based Approach 299
A Coordinate Descent Approach 299
THE CLASSIC SVM PROBLEM 300
GENERAL SEMIPARAMETRIC SVM 301
The Augmented Lagrangian Approach 301
Training Semiparametric SVM 301
A Possible Simplification 302
REGULARIZED PARAMETRIC SVM APPROACH 302
EXPERIMENTAL RESULTS 302
CONCLUSIONS 305
ACKNOWLEDGMENTS 305
REFERENCES 305

CHAPTER 9

Gradient-Based Learning Applied to Document Recognition

INTRODUCTION 306
Learning from Data 307
Gradient-Based Learning 308
Gradient Back-Propogation 308
Learning in Real Handwriting Recognition Systems 309
Globally Trainable Systems 309
CONVOLUTIONAL NEURAL NETWORKS FOR ISOLATED CHARACTER RECOGNITION 310
Convolutional Networks 311
LeNet-5 312
Loss Function 314
CHAPTER 10 Pattern Recognition Using A Family of Design Algorithms Based Upon Generalized Probabilistic Descent Method

ABSTRACT

INTRODUCTION

DISCRIMINATIVE PATTERN CLASSIFICATION

GENERALIZED PROBABILISTIC DESCENT METHOD

DERIVATIVES OF GPD

REFERENCES
## Contents

*Discriminative metric design* 421  
*Minimum error learning subspace method* 422  
Speaker Recognition Using GPD 424  

### CHAPTER 11

**An Approach to Adaptive Classification** 455  
**ABSTRACT** 455  
**INTRODUCTION** 455  
**THE PROPOSED LEARNING METHOD** 456  
- Problem Statement 456  
- Model Inference and Updates 457  
- Restarting 458  
- The Viterbi Algorithm 459  
- Synopsis of Our Learning Method 459  
- Relation to Other Learning Method 459  

#### EXAMPLES 460
- Fixed Linear and Nonlinear Generating Function 460  
- Time-Varying Functions 461  

**CONCLUSIONS AND FUTURE DIRECTIONS** 463  
**REFERENCES** 463  

### CHAPTER 12

**Reduced-Rank Intelligent Signal Processing with Application to Radar** 465  
**INTRODUCTION** 465  
**BACKGROUND** 466  
**KARHUNEN-LOÈVE ANALYSIS** 467  
- The Karhunen-Loève Transformation 467  
- The Karhunen-Loève Expansion 467  
- Implementing the KLT 468  
**THE MULTIPLE SIGNAL MODEL AND WIENER FILTERING** 469  
**THE SIGNAL-DEPENDENT KLT FOR STATISTICAL SIGNAL PROCESSING** 470  
- The KLT and Principal-Components 470  
- The Cross-Spectral Metic: An Intelligent and Signal-Dependent KLT 471  
**INTELLIGENT SIGNAL REPRESENTATION FOR STATISTICAL SIGNAL PROCESSING** 472  
- A New Criterion for Signal Representation and Its Implementation 473  
- A Generalized Joint-Process KLT: Nonunitary Diagonalization of the Covariance Analysis of the JKLT 475  
**RADAR EXAMPLE** 478  
- Radar Signal Processing 478  
- Estimation of the Statistics and Sample Support 479  
- Simulation 480  

ACKNOWLEDGMENTS 432  
REFERENCES 433  
APPENDIX 447  

ABSTRACT 455  
INTRODUCTION 455  
THE PROPOSED LEARNING METHOD 456  
Problem Statement 456  
Model Inference and Updates 457  
Restarting 458  
The Viterbi Algorithm 459  
Synopsis of Our Learning Method 459  
Relation to Other Learning Method 459  
EXAMPLES 460  
Fixed Linear and Nonlinear Generating Function 460  
Time-Varying Functions 461  
CONCLUSIONS AND FUTURE DIRECTIONS 463  
REFERENCES 463  

Reduced-Rank Intelligent Signal Processing with Application to Radar 465  
INTRODUCTION 465  
BACKGROUND 466  
KARHUNEN-LOÈVE ANALYSIS 467  
The Karhunen-Loève Transformation 467  
The Karhunen-Loève Expansion 467  
Implementing the KLT 468  
THE MULTIPLE SIGNAL MODEL AND WIENER FILTERING 469  
THE SIGNAL-DEPENDENT KLT FOR STATISTICAL SIGNAL PROCESSING 470  
The KLT and Principal-Components 470  
The Cross-Spectral Metic: An Intelligent and Signal-Dependent KLT 471  
INTELLIGENT SIGNAL REPRESENTATION FOR STATISTICAL SIGNAL PROCESSING 472  
A New Criterion for Signal Representation and Its Implementation 473  
A Generalized Joint-Process KLT: Nonunitary Diagonalization of the Covariance Analysis of the JKLT 475  
RADAR EXAMPLE 478  
Radar Signal Processing 478  
Estimation of the Statistics and Sample Support 479  
Simulation 480  

xiv
## Chapter 13

**Signal Detection in a Nonstationary Environment Reformulated as an Adaptive Pattern Classification Problem**

### Abstract

484

### Introduction

485

### An Overview of Nonstationary Behavior and Time-Frequency Analysis

488

### Theoretical Background

492

- Multiple Window Estimates
- Spectrum Estimation as an Inverse Problem

### High-Resolution Multiple-Window Spectrograms

499

- Non-Stationary Quadratic-Inverse Theory
- Multiple Window Estimates of the Loeve Spectrum

### Spectrum Analysis of Radar Signals

507

### Modular Learning Machine for Adaptive Signal Detection

517

- How Does the Adaptive Receiver of Fig. 4 Respond to Nonstationary Environment

### Case Study: Radar Target Detection of a Small Target in Sea Clutter

523

- Details of the Receiver
- Detection Results
- Robustness of the Detector

### Cost Functions for Supervised Training of the Pattern Classifiers

532

### Summary and Discussion

535

### Acknowledgments

536

### References

536

## Chapter 14

**Data Representation Using Mixtures of Principal Components**

### Abstract

541

### Introduction

541

### A Spectrum of Representations

542

- Principal Components
- Vector Quantization

### A Mixture of Principal Components

542

### Subspace Pattern Recognition

543

- Similarity Measure
- Class Prototype
- Norm Invariance

### Training

544

- Topological Organization

### Grey-Scale Feature Extraction

546

- Training Data
- Network Basis Vectors
- Segmentation Results
- Illumination Variations
CHAPTER 15  Image Denoising by Sparse Code Shrinkage  554

ABSTRACT  554
INTRODUCTION  554
MAXIMUM LIKELIHOOD DENOISING OF NONGAUSSIAN RANDOM VARIABLES  555
Maximum Likelihood Denoising  555
Modeling Sparse Densities  555
  Laplace Density  555
  Mildly Sparse Densities  556
  Strongly Sparse Densities  557
  Choice of Model  557
  Some Other Models  557
FINDING THE SPARSE CODING TRANSFORMATION  558
MEAN-SQUARE ERROR APPROACH  558
Minimum Mean-Square Estimator in Scalar Case  558
Analysis of Mean-Square Error  559
Minimum Mean Squares Approach to Basis Estimation  559
SPARSE CODE SHRINKAGE  560
COMPARISON WITH WAVELET AND CORING METHODS  560
EXTENSIONS OF THE BASIC THEORY  561
Nongaussian Noise  561
Estimation of Parameters from Noisy Data  561
Non-orthogonal Bases  561
EXPERIMENTS  561
Generation of Image Data  561
Remarks on Image Data  562
  Windowing  562
  The Local Mean  562
  Normalizing the Local Variance  562
Transform Estimation  563
  Methods  563
  Results  563
Component Statistics  564
Denoising Results  565
CONCLUSION  566
REFERENCES  567
APPENDIX  567

INDEX  569

ABOUT THE EDITORS  573