Advances in Experimental Medicine and Biology 823

Changming Sun
Tomasz Bednarz
Tuan D. Pham
Pascal Vallotton
Dadong Wang *Editors*

Signal and Image Analysis for Biomedical and Life Sciences

Springer

Advances in Experimental Medicine and Biology

Volume 823

More information about this series at http://www.springer.com/series/5584

Changming Sun • Tomasz Bednarz • Tuan D. Pham
Pascal Vallotton • Dadong Wang
Editors

Signal and Image Analysis for Biomedical and Life Sciences

Springer

Editors
Changming Sun
Tomasz Bednarz
Pascal Vallotton
Dadong Wang
Digital Productivity Flagship, CSIRO
Sydney, NSW, Australia

Tuan D. Pham
The University of Aizu
Fukushima, Japan

ISSN 0065-2598 ISSN 2214-8019 (electronic)
ISBN 978-3-319-10983-1 ISBN 978-3-319-10984-8 (eBook)
DOI 10.1007/978-3-319-10984-8
Springer Cham Heidelberg New York Dordrecht London

Library of Congress Control Number: 2014955163

Printed on acid-free paper

Springer is part of Springer Science+Business Media (www.springer.com)

Preface

With an emphasis on applications of computational models for solving modern challenging problems in biomedical and life sciences, this book aims to bring collections of articles from biologists, medical/biomedical and health science researchers together with computational scientists to focus on problems at the frontier of biomedical and life sciences. The goals of this book are to build interactions of scientists across several disciplines and to help industrial users apply advanced computational techniques for solving practical biomedical and life science problems.

This book is for users in the fields of biomedical and life sciences who wish to keep abreast with the latest techniques in signal and image analysis. The book presents a detailed description to each of the applications. It can be used by those both at graduate and specialist levels.

We have included 14 chapters in this book. Some of the chapters are extensively revised versions of papers that were presented at the International Symposium on Computational Models for Life Sciences held on 27–29 November 2013 in Sydney, Australia. There are two main parts in the book: signal and image analysis issues within the subjects of the book.

In the first part of the book, Chap. 1 presents a novel visualisation strategy tailored for proteomics data. A dataset is visualised showing phosphorylation events in response to insulin that leads to new insights into the insulin response pathway. A strategy for web-based presentation of data is also described. Chapter 2 proposes a new approach for the modelling of testosterone regulation to identify all model parameters from the hormone concentrations of testosterone and luteinizing hormone. Simulation results are described to reveal behaviour similar to clinical data. Chapter 3 proposes two distinct hybrid algorithms that combine efficient sequential change-point detection procedures with the Cross-Entropy method. Results show effectiveness of the described method. In Chap. 4, two methods for distinguishing between healthy controls and patients diagnosed with Parkinson's disease by means of recorded smooth pursuit eye movements are presented and evaluated. The results are indicative of the potential of the presented methods as diagnosing or staging

tools for Parkinson's disease. Chapter 5 presents an approach for the identification of the Reichardt elementary motion detector model. A pool of spatially distributed elementary motion detectors is considered, and a way of designing the visual stimuli for a certain order of spatial resolution is suggested. Chapter 6 discusses on the complexity ensemble measures for gait time series analysis that could have a significantly wider application scope ranging from diagnostics and early detection of physiological regime change to gait-based biometrics. Chapter 7 presents the development of a motion capturing and load analyzing system for caregivers aiding a patient to sit up in bed. The difference between the performances of the two types of caregivers were found: the professional adopted a posture that was safe and did not stress the lumbar vertebrae, whereas the layperson tended to adopt an unsafe posture. Chapter 8 proposes an unsupervised multi-scale K-means algorithm to distinguish epileptic EEG signals and identify epileptic zones. The experimental results demonstrate that identifying seizure with multi-scale K-means algorithm and delay permutation entropy achieves higher accuracy than that of K-means and support vector machine. Chapter 9 presents a method for the tracking of EEG activity using motion estimation in brain topomaps to understand the mechanism of brain wiring. Authors demonstrate that it is possible to track the path of a signal across various lobes.

In the second part of the book, Chap. 10 presents an approach to processing ultra high-resolution, large-size biomedical imaging data for the purposes of detecting and quantifying vasculature and microvasculature. The results on cerebral and liver vasculatures of a mouse captured at the Shanghai Synchrotron Radiation Facility are presented. Chapter 11 describes a novel way of carrying out image analysis, reconstruction and processing tasks using cloud based service provided on the Australian National eResearch Collaboration Tools and Resources infrastructure. The toolbox is available on the web. Chapter 12 presents an investigation into how Massey University's Pollen Classifynder can accelerate the understanding of pollen and its role in nature. Chapter 13 presents a digital image processing and analysis approach for activated sludge wastewater treatment. Chapter 14 presents a complete system for 3D reconstruction of roots grown in a transparent gel medium or washed and suspended in water.

We thank all the authors for their contributions to this edited book. We also thank Dan Hills and Susan McMaster from CSIRO Contracts and Legal for their help with the Publishing Agreement between Springer and CSIRO. We are grateful to Dr. Thijs van Vlijmen, Sara Germans-Huisman, Magesh Kaarthick Sundaramoorthy, and other editors at Springer and S. Madhuriba at SPi Technologies India Private Ltd. for their help and great support from the beginning to the production of this book. Materials from the American Institute of Physics (AIP) Publishing that are used by some authors are acknowledged and credits are given in the respective chapters within this book.

Sydney, Australia — Changming Sun
Tomasz Bednarz
Aizu, Japan — Tuan D. Pham
Sydney, Australia — Pascal Vallotton
Dadong Wang

July 2014

Contents

Contributors

Yasuko Ando Yokohama City University, Yokohama, Japan

Hiroshi Arisawa Yokohama National University, Yokohama, Japan

Yulia Arzhaeva Digital Productivity Flagship, CSIRO, North Ryde, Sydney, NSW, Australia

Hans Axelson Department of Neuroscience, Neurophysiology, Uppsala University, Uppsala, Sweden

Michael Bain The University of NSW, Sydney, NSW, Australia

Abdullah Bawakid Faculty of Computing and Information Technology, King Abdul Aziz University, Jeddah, Kingdom of Saudi Arabia

Tomasz Bednarz Digital Productivity Flagship, CSIRO, Sydney, NSW, Australia

Leanne Bischof Digital Productivity Flagship, CSIRO, North Ryde, Sydney, NSW, Australia

Neil Burdett Digital Productivity Flagship, CSIRO, Brisbane, QLD, Australia

Jinhai Cai School of Information Technology and Mathematical Sciences, Phenomics and Bioinformatics Research Centre, Australian Centre for Plant Functional Genomics, University of South Australia, Mawson Lakes, SA, Australia

Shiping Chen Digital Productivity Flagship, CSIRO, Sydney, NSW, Australia

Kristina Davis Department of Pathology, University of Michigan, Ann Arbor, MI, USA

Luke Domanski CSIRO IM&T, Sydney, NSW, Australia

Valeriy Gavrishchaka Department of Physics, West Virginia University, Morgantown, WV, USA

Timur Gureyev CSIRO Manufacturing Flagship, Melbourne, VIC, Australia

Simon Haberle School of Culture, History and Language, The Australian National University, Canberra, ACT, Australia

You He Shanghai Synchrotron Radiation Facility (SSRF), Chinese Academy of Sciences, Shanghai Institute of Applied Physics, Shanghai, China

Egi Hidayat Department of Information Technology, Uppsala University, Uppsala, Sweden

Katherine Holt Institute of Natural Resources, Massey University, Palmerston North, New Zealand

Felicitas Hopf School of Culture, History and Language, The Australian National University, Canberra, ACT, Australia

Shoichiro Ito Yokohama National University, Yokohama, Japan

Daniel Jansson Department of Information Technology, Uppsala University, Uppsala, Sweden

Sandeep Kaur Garvan Institute of Medical Research, Sydney, NSW, Australia

The University of NSW, Sydney, NSW, Australia

Muhammad Burhan Khan Faculty of Engineering and Green Technology, Department of Electronic Engineering, Universiti Tunku Abdul Rahman, Kampar, Perak, Malaysia

Alex Khassapov CSIRO IM&T, Melbourne, VIC, Australia

Pankaj Kumar School of Information Technology and Mathematical Sciences, Phenomics and Bioinformatics Research Centre, Australian Centre for Plant Functional Genomics, University of South Australia, Mawson Lakes, SA, Australia

Ryan Lagerstrom Digital Productivity Flagship, CSIRO, North Ryde, Sydney, NSW, Australia

Xue Yong Lee Faculty of Engineering and Green Technology, Department of Electronic Engineering, Universiti Tunku Abdul Rahman, Kampar, Perak, Malaysia

Rongxin Li Digital Productivity Flagship, CSIRO, North Ryde, Sydney, NSW, Australia

Yan Li Faculty of Health, Engineering and Sciences, University of Southern Queensland, Toowoomba, QLD, Australia

David Lovell Digital Productivity Flagship, CSIRO, Canberra, Australia

David K.G. Ma Garvan Institute of Medical Research, Sydney, NSW, Australia

The University of NSW, Sydney, NSW, Australia

Aamir Saeed Malik Department of Electrical and Electronic Engineering, Centre for Intelligent Signal and Imaging Research, Universiti Teknologi PETRONAS, Tronoh, Perak, Malaysia

Per Mattsson Department of Information Technology, Uppsala University, Uppsala, Sweden

Alexander Medvedev Department of Information Technology, Uppsala University, Uppsala, Sweden

Stanley J. Miklavcic School of Information Technology and Mathematical Sciences, Phenomics and Bioinformatics Research Centre, Australian Centre for Plant Functional Genomics, University of South Australia, Mawson Lakes, SA, Australia

Akinobu Nemoto Yokohama City University, Yokohama, Japan

Choon Aun Ng Faculty of Engineering and Green Technology, Department of Environmental Engineering, Universiti Tunku Abdul Rahman, Kampar, Perak, Malaysia

Humaira Nisar Faculty of Engineering and Green Technology, Department of Electronic Engineering, Universiti Tunku Abdul Rahman, Kampar, Perak, Malaysia

Akemi Nomura Yokohama City University, Yokohama, Japan

Karin Nordström Department of Neuroscience, Uppsala University, Uppsala, Sweden

Dag Nyholm Department of Neuroscience, Neurology, Uppsala University, Uppsala, Sweden

Sean I. O'Donoghue Digital Productivity Flagship, CSIRO, Sydney, NSW, Australia

Garvan Institute of Medical Research, Sydney, NSW, Australia

Tuan D. Pham The University of Aizu, Fukushima, Japan

Tatiana Polushina Faculty of Medicine and Dentistry, Department of Clinical Science, University of Bergen, Bergen, Norway

Madawa Priyadarshana Faculty of Science, Department of Statistics, Macquarie University, Sydney, NSW, Australia

Takako Sato Yokohama National University, Yokohama, Japan

Olga Senyukova Department of Computational Mathematics and Cybernetics, Lomonosov Moscow State University, Moscow, Russia

Seong-O Shim Faculty of Computing and Information Technology, King Abdul Aziz University, Jeddah, Kingdom of Saudi Arabia

Georgy Sofronov Faculty of Science, Department of Statistics, Macquarie University, Sydney, NSW, Australia

Christian Stolte Digital Productivity Flagship, CSIRO, Sydney, NSW, Australia

Ahmad Rauf Subhani Department of Electrical and Electronic Engineering, Centre for Intelligent Signal and Imaging Research, Universiti Teknologi PETRONAS, Tronoh, Perak, Malaysia

Changming Sun Digital Productivity Flagship, CSIRO, North Ryde, Sydney, NSW, Australia

Piotr Szul Digital Productivity Flagship, CSIRO, Sydney, NSW, Australia

Yosuke Takami Yokohama National University, Yokohama, Japan

Hai Tan Shanghai Synchrotron Radiation Facility (SSRF), Chinese Academy of Sciences, Shanghai Institute of Applied Physics, Shanghai, China

John A. Taylor Digital Productivity Flagship, CSIRO, Canberra, ACT, Australia

Darren Thompson CSIRO IM&T, Melbourne, VIC, Australia

Rafi Ullah Comsats Institute of Information Techonology, Islamabad, Pakistan

Pascal Vallotton Digital Productivity Flagship, CSIRO, Sydney, NSW, Australia

Dadong Wang Digital Productivity Flagship, CSIRO, North Ryde, Sydney, NSW, Australia

Shuaifang Wang Faculty of Health, Engineering and Sciences, University of Southern Queensland, Toowoomba, QLD, Australia

Peng (Paul) Wen Faculty of Health, Engineering and Sciences, University of Southern Queensland, Toowoomba, QLD, Australia

Tiqiao Xiao Shanghai Synchrotron Radiation Facility (SSRF), Chinese Academy of Sciences, Shanghai Institute of Applied Physics, Shanghai, China

Tomohiro Yano Yokohama National University, Yokohama, Japan

Kim Ho Yeap Faculty of Engineering and Green Technology, Department of Electronic Engineering, Universiti Tunku Abdul Rahman, Kampar, Perak, Malaysia

Guohun Zhu Faculty of Health, Engineering and Sciences, University of Southern Queensland, Toowoomba, QLD, Australia

School of Electronic Engineering and Automation, Guilin University of Electronic Technology, Guilin, China

Acronyms

aCGH	Array comparative genomic hybridization
ALS	Amyotrophic lateral sclerosis
ANN	Artificial neural networks
ARC	ANU reference collection
ARL	Average run length
AS	Activated sludge
AU	Area under
BBME	Block based motion estimation
BMA	Block matching algorithms
BME	Block motion estimation
BOD	Biological oxygen demand
CBS	Circular binary segmentation
CE	Cross-entropy
CGH	Comparative genomic hybridization
CLSM	Confocal laser scanning microscopy
CMR	Central motor region
CMVs	Candidate motion vectors
CNV	Copy number variation
COD	Chemical oxygen demand
CT	Computed tomography
CTE	Cortical thickness estimation
CTS	Classifynder test set
CUSUM	Cumulative sum
DFA	Detrended fluctuation analysis
DO	Dissolved oxygen
DPE	Delay permutation entropy
DREAM	Dialogue for Reverse Engineering Assessments and Methods
DS	Diamond search
DT	Decision trees

ECG	Electrocardiography
EDL	Ensemble decomposition learning
EEG	Electroencephalogram
EMD	Elementary motion detector
EOG	Electrooculography
EZ	Epileptogenic zone
FBP	Filtered back-projection
FDK	Feldkamp-Davis-Kress
FFT	Fast Fourier transform
FS	Full search
FSS	Four step search
GFT	Grass fire transform
GGCM	Grey gradient co-occurrence matrix
GMM	Gaussian mixture model
GnRH	Gonadotropin-releasing hormone
HD	Huntington's disease
HMMs	Hidden Markov models
HRT	Hydraulic retention time
HRV	Heart rate variability
HS	Horizontal system
IaaS	Infrastructure as a service
IBL	Instance-based learning
ICA	Imaged absolute conics
iEEG	Intracranial EEG
KDE	Kernel density estimation
LDA	Linear discriminant analysis
LDSP	Large diamond search pattern
LH	Luteinizing hormone
LPTCs	Lobula plate tangential cells
LTR	Left temporal region
MAD	Mean absolute difference
MCRT	Mean cell residence time
MFA	Multi-fractal analysis
MGD	Million gallons per day
MIMO	Multiple-input multiple-output
MRI	Magnetic resonance imaging
MSE	Mean square error
MSEn	Multi-scale entropy
MST	Minimum spanning tree
MV	Motion vectors
NIOSH	National Institute for Occupational Safety and Health
NLD	Nonlinear dynamics
NN	Neural networks
OSA	Orthogonal series approximation
OSAlg	Orthogonal search algorithm

PaaS	Platform as a service
PAR	Peak-to-average-ratio
PD	Parkinson's disease
PDB	Protein database
PDF	Probability density function
PE	Permutation entropy
PELT	Pruned exact linear time
PET	Positron emission tomography
PRA	Pel-recursive algorithms
PSNR	Peak signal-to-noise ratio
PVC	Partial volume correction
RAS	Return activated sludge
RBF	Radial basis function
REM	Rapid eye movement
RF	Random forests
RMSE	Root mean square error
RSA	Root system architecture
RTR	Right temporal region
SaaS	Software as a service
SAD	Sum of absolute differences
SC	Stopping criterion
SDD	Sample-to-detector distance
SDI	Sludge density index
SDSP	Small diamond search pattern
SE	Sample entropy
SEL	Single-example learning
SIFT	Scale-invariant feature transform
SISO	Single-input single-output
SNR	Signal-to-noise ratio
SPEM	Smooth pursuit eye movements
SPG	Smooth pursuit gain
SPS	Smooth pursuit system
SR	Shiryaev-Roberts
SR-μCT	Synchrotron radiation based micro-computed tomography
SSIM	Structural similarity index measure
SSRF	Shanghai Synchrotron Radiation Facility
SUVR	Standard uptake value ratio
SVI	Sludge volume index
SVM	Support vector machine
TDLS	Two-dimensional logarithmic search
TDR	True detection rate
Te	Testosterone
TIE	Transport of intensity equation
TOC	Total organic carbon
TS	Total solids

TSS	Three step search
TSSol	Total suspended solids
UESA	Unimodal error surface assumption
UPDRS	Unified Parkinson's Disease Rating Scale
VG	Visibility graph
VSS	Volatile suspended solid
WWTP	Wastewater treatment plants
ZMs	Zernike moments

Part I
Signal Analysis

Chapter 1
Visual Analytics of Signalling Pathways Using Time Profiles

David K.G. Ma, Christian Stolte, Sandeep Kaur, Michael Bain, and Seán I. O'Donoghue

Abstract Data visualisation is usually a crucial first step in analysing and exploring large-scale complex data. The visualisation of proteomics time-course data on post-translational modifications presents a particular challenge that is largely unmet by existing tools and methods. To this end, we present Minardo, a novel visualisation strategy tailored for such proteomics data, in which data layout is driven by both cellular topology and temporal order. In this work, we utilised the Minardo strategy to visualise a dataset showing phosphorylation events in response to insulin. We evaluated the visualisation together with experts in diabetes and obesity, which led to new insights into the insulin response pathway. Based on this success, we outline how this layout strategy could be automated into a web-based tool for visualising a broad range of proteomics time-course data. We also discuss how the approach could be extended to include protein 3D structure information, as well as higher dimensional data, such as a range of experimental conditions. We also discuss our entry of Minardo in the international DREAM8 competition.

Keywords Visual analytics • Signalling pathways • Proteomics • Temporal data • Graph layout • Phosphorylation • Insulin response

D.K.G. Ma • S. Kaur
Garvan Institute of Medical Research, Sydney, NSW, Australia

The University of NSW, Sydney, NSW, Australia
e-mail: davidma@cse.unsw.edu.au; sandeep.kaur@unsw.edu.au

C. Stolte
Digital Productivity Flagship, CSIRO, Sydney, NSW, Australia
e-mail: christian.stolte@csiro.au

M. Bain
The University of NSW, Sydney, NSW, Australia
e-mail: mike@cse.unsw.edu.au

S.I. O'Donoghue (✉)
Digital Productivity Flagship, CSIRO, Sydney, NSW, Australia

Garvan Institute of Medical Research, Sydney, NSW, Australia
e-mail: sean@odonoghuelab.org

C. Sun et al. (eds.), *Signal and Image Analysis for Biomedical and Life Sciences*,
Advances in Experimental Medicine and Biology 823,
DOI 10.1007/978-3-319-10984-8_1

1.1 Introduction

Computationally aided data visualisation is helpful for analysing and exploring large-scale complex data as it allows computational abilities, such as large memory capacities and fast calculations, to be combined with human abilities, such as high-bandwidth visual perception and creativity, to address the task of understanding such data [19]. With the emergence of large-scale and high-dimensional datasets in molecular systems biology, the task of data visualisation has become increasingly important [28].

Current high-throughput technologies typically enable thousands of molecules to be tracked simultaneously. One such high-throughput method uses mass spectrometry to enable the quantification of the phosphorylation state of each protein in a cell's proteome. In typical experiments of this type, cells are initially stimulated with an agent (e.g., insulin, glucose, or a range of inhibitor molecules) and the response is measured at discrete points in time. The temporal order of such time-series experiments offers great potential to prioritise paths in the resulting dense protein interaction graphs [11].

In order to understand biomolecular systems it is essential to understand how the interactions of their component molecules result in the overall changes in cell physiology – for example, how a fat cell initially starved of glucose switches to active uptake and processing of glucose upon stimulation by insulin. The most common approach used to gain an understanding of such events is to draw graphs of signalling pathways [14]. These pathway maps definitely have their limitations: for example, as explained by Kitano [20], they could be thought of as analogous to static road maps, when what we really wish to know are the traffic patterns, why such patterns emerge and how we can control them. Nonetheless, visualisations of pathway maps are an important first step.

There are several initiatives worldwide aimed at consolidating all human knowledge about biological systems into a single, searchable database and with the results presented in the form of interactive pathways graphs. Currently however there is no consensus about a single 'best' approach – instead, there are a large number of different databases, each with a tailored visualisation system. Some of the more widely used resources of this kind include Pathway Commons [6], KEGG [18], PANTHER [25], BIOCARTA [27], and Reactome [24].

When considering new data from high-throughput experiments, a common strategy is to visually overlay these data onto existing pathway graphs extracted from one of the above resources. A wide variety of methods and tools have been developed to facilitate overlaying experimental data onto pathways, including Pajek [3], BiologicalNetworks [1], Medusa [31], as well as many plug-ins to the Cytoscape framework [36]. A recent review of such methods for 'omics' data is provided by Gehlenborg et al. [14]. However, as noted in [14], the major challenge for visualisation methods is how to benefit from the explosion in dataset scale and complexity without overwhelming the user. This is a difficult problem which

currently has no obvious general solution, but we suggest the answer should lie in how *context* may be used in visualisation. The contribution of this paper lies in the adoption of a novel visual metaphor that can illustrate significant temporal and potentially causal relationships in high-throughput data on cell signalling pathways.

1.1.1 Challenges in Visualising High-Throughput Time-Series Post-translationally Modified Proteomic Datasets

The biology of the post-translational modification of proteins presents some important issues for visualisation. Firstly, there are many different types of such modifications that we may require to be visualised (e.g., phosphorylation, methylation, sumoylation, etc.). Since different modifications are typically implicated in different functional roles, indications of these differences could be critical for successful visualisation.

Secondly, most of the current network-based visualisation tools for high-throughput datasets have been designed for gene expression. However, it is not always possible to simply reuse such tools for proteomics datasets that incorporate post-translational modifications. For example, when viewing time-series transcriptomic datasets, we are usually interested in the expression levels of *whole RNA molecules* over time – for such data, time-profile information is often added to a network view by adding colouring or a pattern to each node or edge [14, 35]. However for proteomics datasets showing post-translational modifications a more detailed representation is required, since we are typically interested in the abundance levels of *multiple residues* within each protein.

As an example, we recently conducted a pilot user study [21] to evaluate the reuse of the Cytoscape plug-in 'Cerebral' [2] to visualise the proteomic dataset of Humphrey et al. [15]. Cerebral was initially designed for use with gene expression data – we found that although several aspects of this tool were of benefit, overall the layout and representation concepts of the tool were not well suited to visualising post-translational modifications.

A general problem faced by omics visualisation tools is the challenge of facilitating simultaneous visualisation of multiple kinds of experimental data. For example, how can high-throughput time-series data on post-translational modifications be visualised in an coherent and integrated way with other data, such as abundance level for transcripts or protein?

Evidence of the growing recognition of the need and importance of this type of integrated omics visualisation comes in the form of the latest iteration of the international DREAM (Dialogue for Reverse Engineering Assessments and Methods) competition. The DREAM8 Sub-challenge 3: "Visualisation of high-dimensional time-course on breast cancer proteomics data" was designed to facilitate research on novel tools for this purpose.

1.1.2 Aims

In this paper we outline the elements of the Minardo visualisation concept (Fig. 1.1) that we developed recently to address the above challenges – Minardo is based on using cell topology combined with temporal ordering as the key layout contexts used to organize how the data is depicted. In this study, we worked with an experimental research group that is applying state-of-the-art methods in high-throughput experimental proteomics to study the time course of protein phosphorylation events in human cells in vitro following stimulation by insulin [15], as part of a broader project on diabetes and obesity. The group had already applied a wide range of existing analysis and visualisation tools to these data, although relatively few tools were specifically tailored for time-course phosphophorylation data. The group's key unmet requirement was for a system that would enable visual exploration of networks representing insulin response, which could be interactively overlaid firstly with phosphophorylation time-course data, and that later could also include data on RNA and protein abundance.

To address this need, we first tried several existing visual analytics approaches with the goal of representing the data to gain new insight. From discussion of the merits and weaknesses of these existing approaches with our experimental collaborators we used this feedback, together with visual analytics principles, to develop an improved general layout strategy specifically for time-series post-translationally modified proteomic data.

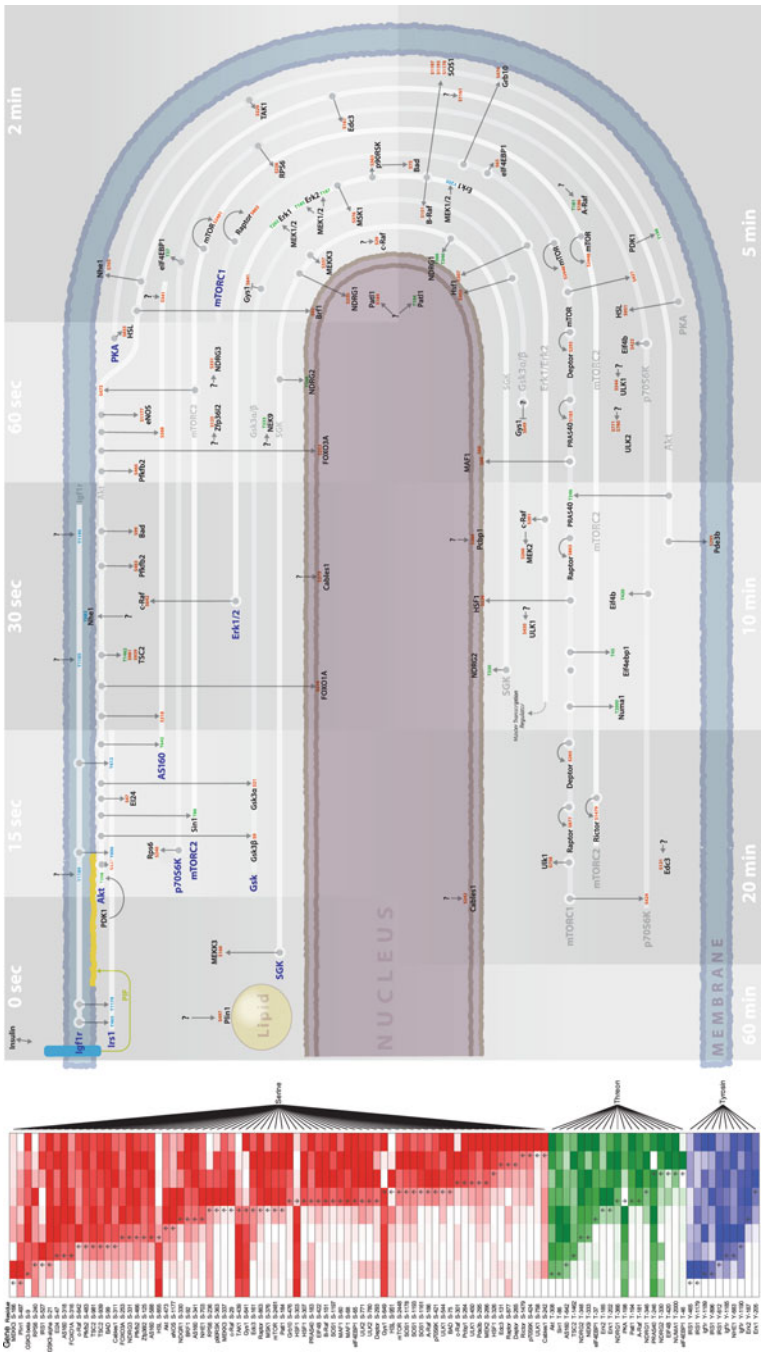
We called our layout strategy "Minardo" as a play on words, as the layout was partly inspired by the well-known information graphic published by Minard in 1869.[1] A key innovation that comes from this inspiration is the ability to combine aspects of network-based structure with temporally ordered event profiles.

As discussed in [21], the Minardo approach has proved helpful, having revealed several inconsistencies with the previously published interpretation of this dataset, and suggested several new insights into the timing and order of events underlying the insulin response pathway.

While the current layout has been constructed specifically for analysing phosphorylation data related to insulin response, aspects of the layout have clear potential to be generalised to help with analysing a broader range of systems biology data. Thus, we are doing ongoing work aimed at developing the Minardo layout strategy into a general tool.

The remainder of this paper is structured as follows. In Sect. 1.2 we describe how to create the Minardo layout and how to connect it interactively with a heat map visualisation of the same data. Section 1.3 presents results from a user study to evaluate Minardo, and our entry using Minardo into the DREAM8 visualisation challenge. In Sect. 1.4 we discuss implications of this approach and outline directions for future work. Section "Conclusions" concludes the paper.

[1]This famous graphic shows Napoleon's disastrous Russian campaign of 1812 – the graphic is regarded as an exemplar by many data visualisation specialists [4].


0 sec
15 sec
30 sec
60 sec
2 min
5 min
10 min
20 min
60 min
NUCLEUS
MEMBRANE
Insulin
Igf1r
Irs1
Lipid
PDK1
Akt
PKA
mTORC1
mTORC2
Erk1/2
AS160
Gsk
SGK
Raptor
Rictor
Deptor
Sin1
FOXO1A
FOXO3A
Cables1
Bad
Bim
Bcl2
Bcl2l1
Bax
Bak1
Mcl1
Noxa1
Puma
Bid
Bmf
Hrk
Bik
Fas
Fasl
Tnfrsf10b
Tnfsf10
Tnfrsf1a
Tnf
Trail
Tnfrsf10a
Tnfrsf10c
Tnfrsf10d
Tnfrsf11b
Tnfrsf12a
Tnfrsf21
Tnfrsf25
Tnfsf10
Tnfsf11
Tnfsf12
Tnfsf13
Tnfsf13b
Tnfsf14
Tnfsf15
Tnfsf18
Tnfsf8
Tnfsf9
Xiap
Birc2
Birc3
Birc5
Birc6
Dffa
Dffb
Diablo
Omi
Casp1
Casp2
Casp3
Casp4
Casp6
Casp7
Casp8
Casp9
Casp10
Casp12
Casp14
Cflar
Fadd
Ripk1
Ripk3
Trp53
Cdkn1a
Cdkn1b
Cdkn2a
Cdkn2b
Cdkn2c
Cdkn2d
Gadd45a
Gadd45b
Gadd45g
Mdm2
Mdm4
Atm
Atr
Chek1
Chek2
Rad9a
Rad17
Rad1
Hus1
Rad51
Rad51b
Rad51c
Rad51d
Rad52
Rad54l
Rad54b
Brca1
Brca2
Brip1
Palb2
Rbbp8
Blm
Wrn
Recql
Recql4
Recql5
Xrcc1
Xrcc2
Xrcc3
Xrcc4
Xrcc5
Xrcc6
Ercc1
Ercc2
Ercc3
Ercc4
Ercc5
Ercc6
Ercc8
Ddb1
Ddb2
Rnf8
Rnf168
Rnf169
Rnf20
Rnf40
Ube2a
Ube2b
Ube2n
Ube2v1
Ube2v2
Ube2d1
Ube2d2
Ube2d3
Ube2e1
Ube2e2
Ube2e3
Ube2g1
Ube2g2
Ube2i
Ube2k
Ube2l3
Ube2m
Ube2q1
Ube2q2
Ube2r2
Ube2s
Ube2t
Ube2w
Ube2z
Ube2j1
Ube2j2
Ube2c
Ube2h
Ube2f
Ube2d4
Ube2d5
Ube2d6
Ube2d7
Ube2d8
Ube2d9
Ube2d10
Ube2d11
Ube2d12
Ube2d13
Ube2d14
Ube2d15
Ube2d16
Ube2d17
Ube2d18
Ube2d19
Ube2d20
Ube2d21
Ube2d22
Ube2d23
Ube2d24
Ube2d25
Ube2d26
Ube2d27
Ube2d28
Ube2d29
Ube2d30
Ube2d31
Ube2d32
Ube2d33
Ube2d34
Ube2d35
Ube2d36
Ube2d37
Ube2d38
Ube2d39
Ube2d40
Ube2d41
Ube2d42
Ube2d43

1.2 Methods

Our visualisation strategy consists of two main components – the Minardo layout, and a heat map – both of which are connected to support interactive data exploration. This is shown in Fig. 1.1, using the insulin response dataset of Humphrey et al. [15].

The Minardo layout depicts a cellular topology, divided into regions that represent the time points of the time course data. Note that although time points typically denote discrete values fixed by the experimental protocol (e.g., 0, 15, 30 s, etc.), the Minardo layout allows placement of events ordered along a continuous time scale in cases where continuous data are available (e.g., either directly from experiments that measure continuous time values, or perhaps estimated by interpolation from discrete data). The tracks across time points indicate individual proteins or protein complexes that are active (in terms of events in the dataset) over multiple time points. Since screen real estate is limited, only a limited number of tracks can be displayed at the same time. In many cases the number of tracks that can be shown will be insufficient to show all proteins active across multiple time points – thus, some criteria will need to be applied to select which proteins are to be displayed on the available tracks. A natural criterion is the level of activity of a protein, as this suggests its importance, although other criteria could also be applied.

Causal relationships between different proteins within a time point are depicted using directed edges running perpendicular to the tracks; for the insulin response dataset [15], these relationships link a kinase to its phosphorylation substrate. The actual protein residue number of the phosphorylated amino acid – known as the *phosphosite* – is shown colour-coded in Fig. 1.1.

In the rest of this section we describe in further detail the method and the design decisions used to create the layout. We also describe the heat map, and the procedures used to link the layout to the heat map, as well as the implementation procedures used to make the visualisation interactive.

1.2.1 Phosphorylation Dataset for Insulin Response

We worked with members of the James laboratory[2] at the Garvan Institute of Medical Research, a world-leading laboratory in applying experimental systems biology to study diabetes and obesity. They recently published a study of the time course of protein phosphorylation events occurring in vitro in mouse 3T3-L1 adipocyte cells – cells derived from brown adipose (fat) tissue – after 'feeding' (i.e., stimulation by insulin and glucose) [15].

The cell used in the experiment was initially in a starved state, then stimulated with insulin and glucose. The cells were lysed at 0, 15, 30 s and 1, 2, 5, 10, 20 and 60 min after stimulation. Mass spectrometry was then used to measure the

[2] http://www.jameslab.com.au

phosphorylation state of all detectable Serine (S), Threonine (T) and Tyrosine (Y) amino acid residues [9], resulting in a final set of time profiles for 7,897 phosphosites that were judged to be of good quality – an average of about 6.5 phosphosites per protein [15].

Humphrey et al. then used unsupervised fuzzy C-means clustering to organise the time profiles for each phosphosite into groups [15]. They also conducted an extensive literature survey to identify the kinases responsible for a subset consisting of 104 of the phosphosites judged to be most significant, based on prior knowledge of the response pathway. These data – presented in Fig. 5 of Humphrey et al. [15] – were used as the starting point for our work, with the goal of re-analysing and organising these data to provide greater insight into underlying biological processes.

1.2.1.1 Data Representation

The phosphorylation time-series data from Humphrey et al. [15] were generated by the MaxQuant software [10]. We obtained these data as a comma-separated text file, which contained the ratios of the absolute values of observed levels of phosphorylation for each phosphosite at each time-point to a basal level. The basal level represented the phosphorylation levels in starved cells, and the time-points represented the phosphorylation levels after that amount of time has elapsed since stimulation with insulin and glucose. The dataset consisted of triplicate measurements of phosphorylation levels for each of the nine time-points. Phosphorylation levels at time zero were set to 1.0, and the phosphorylation level at each subsequent time point was the ratio of that point's abundance to its basal level (for more information, see the Methods section of Humphrey et al. [15]).

1.2.2 Heat Map of the Time-Series Data

To display the complete time-course data, we used a traditional heat map. We utilised three colour scales, red, green, and blue to represent *Serine*, *Threonine* and *Tyrosine* residue phosphosites, respectively. The heat map depicted only those profiles which were also present in the Minardo layout.

In order to create the heat map, we averaged the triplicate values at each time-point, then we linearly rescaled the resulting time profile, setting the lowest level of activation achieved across the time-series to 0 % and the highest level to 100 %. Finally, we used the JavaScript library D3.js [5] to create an interactive heat map visualisation of these data.

1.2.2.1 Selecting a Single Time Point for Each Phosphorylation

Using the re-scaled data, we devised a method for consistently selecting a single representative time-point for each phosphosite. Based on an analogy to the Michaelis

constant [26] in enzyme kinetics, we estimated the time at which each phosphosite first transitions from either below its 50 % level to above, or vice versa in the case of a dephosphorylation event. We took this to be the first activation time, and marked it on the heat map using either an up or down arrow to indicate phosphorylation or dephosphorylation, respectively.

This induces a linear (total) ordering on the data, where each phosphosite's time course is denoted by its estimated first activation time. Note also that the activation times estimated by this method have continuous values, hence effectively increasing the temporal resolution of the dataset, which is also useful for constructing the Minardo layout.

1.2.3 The Minardo Layout

The Minardo layout was constructed using a number of graphic design principles, combined with user feedback, and drawing from concepts used in existing tools, such as the Cerebral plug-in [2]. The visual channels used – primarily position, hue, and connection – were chosen to effectively convey key information with low cognitive load [7, 38].

Position is usually the most powerful visual channel [22] hence in Minardo we have used the X and Y axis to show time and sub-cellular topology, as they are key features of the dataset. We created a schematic cell in Adobe Illustrator, mapping time in an arc around the cell and adding intervals to represent the time points used to derive the experimental data (Fig. 1.1). With a single first activation time identified for each phosphosite, phosphorylation events could be placed unambiguously within one specific time interval on the diagram. We also arranged the cell topology such that the regions for each time interval contain extracellular space, cytoplasm, and nuclear space, allowing for positioning proteins based on their subcellular location.

Rather than laying out the consecutive time intervals in one direction (e.g., along the X or Y axes), we have taken inspiration from Charles Joseph Minard's classic flow map of Napoleon's March [4] and wrapped the flow of time around the cell topology, creating an overall aspect ratio that allows the entire diagram to more easily fit the landscape orientation of most computer displays. Wrapping time in this way also allow connections from later to earlier time points, providing clear representation of feedback loops.

Lines with arrows were used to indicate kinases and their target phosphosites. In the current dataset there are 104 such connections. To overcome the typical 'hairball' problem that occurs with networks of this size and larger, we reduced clutter by using tracks to represent 'promiscuous' proteins or complexes, i.e., those involved in multiple phosphorylation events at multiple time-points. This is similar to the concept of hubs, or high-degree nodes of a network, but modified to account for the time-series dataset.

Hue was used consistently in the network and heat map, with red, green, and blue used to represent Serine, Threonine and Tyrosine residues. Yellow was used to highlight items selected by the user. The default highlight was Yellow only,

and it showed the relevant kinases and phosphorylation events on the track, or the phosphates currently being brushed over. “Show Targets” is a toggle button, which turns Teal when switched on, indicating to the user that phospho-targets are now being shown with a Teal highlight.

The layout was saved in SVG format and imported in an HTML page with the heat map. JavaScript was used to implement brushing and linking between the two representations.

1.3 Results

Our implementation of the Minardo layout and heat map applied to the insulin response phosphorylation time course data [15] resulted in a single HTML file, which is included in the supplementary information (http://odonoghuelab.org/Minardo.zip). A screen shot of this HTML file can be seen in Fig. 1.1. Two distinct components are clearly seen in this figure, the heat map and the Minardo Layout. The HTML files supports interactivity between these components via brushing and linking. For example, hovering over a protein (in either the heat map or the Minardo layout) automatically highlights all occurrences of the protein name in the HTML document. Text searching of the HTML document can also be done, using standard browser functionality, resulting in highlighting of all proteins with names that match the search term.

In the Minardo layout, the insulin response network (taken from Fig. 5 of Humphrey et al. [15]) has been overlaid on a typical cellular topology. This cellular topology has been divided into a number of time-points as present in the dataset – in this case, nine time points. It shows the temporal order of phosphorylation events, with arrows identifying each kinase and, its substrate phosphosite. The proteins Akt, Irs1, AS160, p70S6K, Erk1 and Erk2, and the complexes Gsk, mTORC1 and mTORC2, play roles across multiple times and so have been indicated with white tracks running parallel to the cellular membrane. For each of the protein phosphorylation sites featured in the Minardo layout, an entry has been created in the interactive heat map, showing its normalised abundance levels detected across each of the time points. The HTML allows sorting the heat map in multiple ways including by residue type, by UniProt identifier [23], by identified time of first regulation, and many more.

1.3.1 Evaluation of the Minardo Visualisation Strategy

We conducted an informal user study with experts in the field of diabetes and obesity studying insulin response at the Garvan Institute [17]. During the study, the interactive HTML file was disseminated to the users by making it available within the organisation’s intranet. Users were asked to freely explore the use of the visualisation strategy and provide feedback.

To summarise the results of this study, the users judged the Minardo visualisation favourably. They found the brushing and linking feature between the heat map and the network to be very helpful for interpreting the data in detail. The most positive feedback, however, was that the new layout helped them gain new insights into the underlying bio-molecular processes. These new insights are detailed in our related work in Ma et al. [21]. The validations of these insights are underway, with a joint publication with the biologists in preparation.

1.3.1.1 Requested Features

In this study, the users requested a number of features not yet supported by our current Minardo implementation.

First and foremost was the ability to easily select which phosphorylation sites are used to construct the layout. The current dataset shows only 104 of the 7,897 high quality phosphorylation sites that were present in the original dataset – these 104 sites were selected by Humphrey et al. [15] as they were believed to be the most important. Nonetheless, the users would like the facility to examine other subsets of phosphosites using the Minardo layout.

A second requested feature was the ability to interactively edit the network in order to change the assignment of kinases to targets. A third requested feature was the ability to add additional data (or datasets), such as multiple experimental conditions or the presence of various chemical inhibitors. Finally, users requested that the facility to search proteins by name be extended so as to match different synonyms for the same protein – this would be very useful since many proteins used in this study are known by multiple names (e.g., As160, Kiaa0603, Tbc1d4 all refer to the same protein).

1.3.2 Minardo in the International DREAM8 Competition

The aim of the DREAM8 'visualisation of high-dimensional time-course on breast cancer proteomics data' sub-challenge was to propose novel strategies to visualise high-dimensional molecular time-course data. The datasets provided for the competition featured phosphorylation proteomics data, for approximately 45 phosphosites, at seven time points, under eight stimulus conditions. Data was also given for a control, and under conditions in which the phosphorylation ability of 3 crucial kinases was inhibited.

We entered Minardo in this competition. All of its features as described above were presented, however additional modifications were proposed to its workflow to enable comparison between multiple different experimental conditions (stimuli, inhibitor or control). Figures 1.2 and 1.3 show the main visualisations in the proposed Minardo workflow when visualising a single set of conditions and when comparing between two conditions.

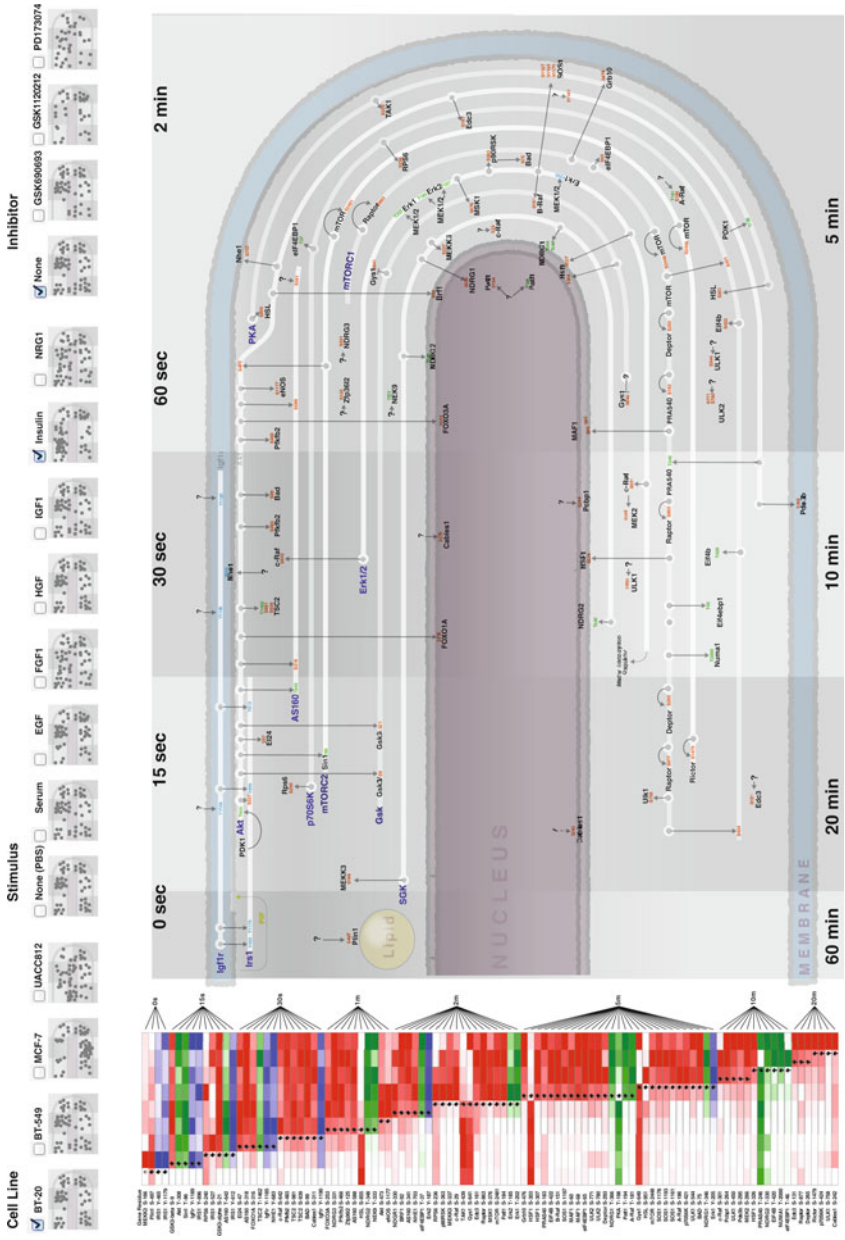


Fig. 1.2 Visualising a single set of conditions as selected interactively by the user. Additional checkboxes are placed above the Minardo layout to enable the user to select from a set of conditions (in this case, to pick a Cell Line, a Stimulus, and Inhibitor). The rest of the visualisation is as before

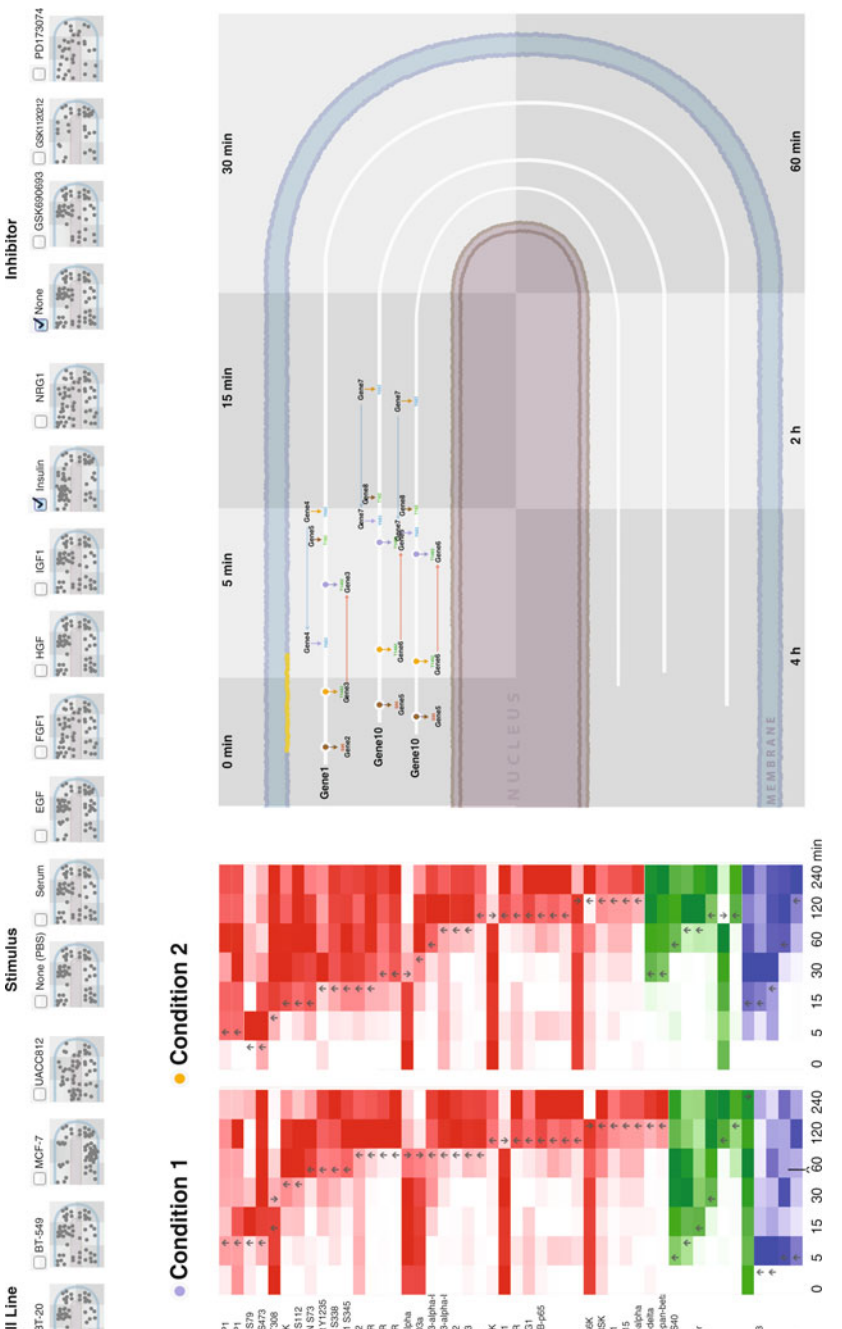


Fig. 1.3 Minardo visualisation strategy for comparing two experimental conditions. See text for details

1.3.2.1 Proposed Workflow

We envisioned three visualisation scenarios: visualising a single set of conditions (cell line, stimulus and inhibitor or control), visually comparing two such sets of conditions, and comparing three or more sets of conditions.

When visualising a *single set of conditions*, the Minardo visualisation strategy remains unchanged, that is, it is as we have described it in this paper so far. Figure 1.2 shows this scenario for the competition data, with the only addition the 'check-boxes' which allow selection of any particular condition.

When *two conditions are selected*, we expect that key change between the two conditions would be the difference in the timings of phosphorylation events. As such, to visualise this, we propose checkboxes that allow the user to either 'Show differences', 'Show similarities', or 'Show all'. This could modify the nodes in the main network to selectively highlight the nodes responding differently under the two conditions. As such, each phosphorylation event could be represented by two nodes, indicating the shift in timing for each event between the different conditions, with the nodes connected by track which includes an arrowhead indicating the direction of change in timing. Alternatively, node colour could also be used to indicate the time shift from earlier to later events. For example, red could be used to indicate nodes having moved earlier in time, and blue could be used to indicate nodes having moved later in time (right of Fig. 1.3).

When visualising two conditions, we also propose displaying two sets of heat maps corresponding to the two conditions, where the rows corresponding to a particular phosphorylation site are aligned next to each other to enable easy comparison. Figure 1.3 shows the proposed workflow for this scenario.

Finally, when *three or more conditions are selected*, our entry proposed coupling the Minardo layout with a standard dimension reduction method, such as principal component analysis, that can provide an abstract, high-level overview of how the conditions compare, and – by brushing and linking – allow the user to easily drill down to investigate either single conditions, or compare two sets of conditions.

The DREAM8 competition was useful in providing a realistic set of multidimensional experimental data. These data provided the impetus for us to extend our plans for Minardo to accommodate such a scenario. Our entry was judged positively by other entrants in the DREAM8 competition, with an overall rank of 6 out of the 14 total submissions. However, to our knowledge the Minardo layout was the only novel visualisation metaphor proposed amongst the DREAM8 submissions.

1.4 Discussion and Further Work

The Minardo layout is a novel visual analytics approach for time-series data, such as phosphorylation response profiles – we are continuing to test and extend this approach. Thus far, our first implementation of Minardo layout strategy has proven effective in providing a framework for re-interpreting the time course of

phosphorylation events following insulin stimulation. Minardo specifically includes protein sub-cellular localisation as part of the layout, and we have found this greatly aided in interpreting the biological significance of the data.

A key principle behind the Minardo layout is that it enables visualising *time* along with changes in the network. This allows for spatial and temporal reasoning about the separate stages of the experiment, thus effectively subdividing the entire dataset into more manageable chunks.

A second key element of the visualisation is that promiscuous proteins and complexes – i.e., those involved in multiple phosphorylation events at multiple time-points – are represented as tracks running parallel to the cellular membrane and spanning multiple time intervals. These tracks are represented in a form that is visually very distinct from the edges that represent phosphorylation events, with the result that this makes these important hub proteins highly visible. This also results in a graph with no apparent edge crossing, as can be seen in Fig. 1.1, a very beneficial outcome for helping users understand and interpret information in graphs [8]. The use of such 'tracks' has some similarity with the use of lines in a parallel co-ordinate plot [16], which represents a single point of information that has values across multiple axes.

The third key principle is the use of the 'first activation time', which is used to map each phosphorylation site to a single time point on the layout. This is a significant simplification compared to the full time course data, which greatly helps users in interpreting the data. However, when needed the complete time course data is easily accessible to the user; it is visualised in the heat map, and is connected to the Minardo layout by interactive brushing and linking.

1.4.1 Minardo as a Web-Based Tool

In its current state, the Minardo layout has been created by searching various resources and gathering information on the various components, such as the network components, the sub-cellular localisations of phosphorylation events and the proteins, individually, to manually assemble the visualisation. However, in order for it to be a tool reusable for visualising data on various different bio-molecular systems, many of its tasks will require automation. We envisage several key steps will be required to automate this:

1. Selection of most significant phosphosites;
2. Identification of potential kinases for each phosphorylation events;
3. Sub-cellular localisation of proteins and phosphorylation events.

For selecting the most significant phosphorylation events, several metrics could be developed to analyse each experimental dataset and find the phosphosites undergoing the most significant changes in phosphorylation state (e.g., identifying the sites showing most variation across the time course of the experiment). In addition, these metrics could be supplemented with knowledge derived from common

pathway and network database resources such as KEGG, Pathway Commons, BIOCARTA and NetworKIN. For the dataset used in this study, this knowledge was used to filter the total of nearly 8,000 sites down to just 104 – a key next step for Minardo would be to automated this process, for example, using scoring functions to rank proteins in order of importance, based on connectivity in previously determined networks and pathways.

For the identification of potential kinases, we propose the use of the above common pathway resources, as well as resources more specifically focused on phosphorylation data, such as Phospho.ELM [13].

Finally, for determining the sub-cellular localisation of each phospho-event, we propose the use of specialist cellular localisation databases, such as LOCATE[3] or COMPARTMENTS [32], that consolidate observed and predicted location information for large numbers of proteins.

Automation of these steps would enable Minardo to be extended into a reusable web-based tool, where users can upload, high throughput data, and visualise the data in 'context', as well as perform analysis.

1.4.2 Lessons from the Usability Study

Taking into account the feedback obtained from the pilot user study in this work, we plan to extend Minardo, making it more interactive – in particular, to enable user selection of phospho-events – as well as introducing editability, allowing users to reassign the connections between kinases and their target phosphoproteins. However, depending on how these functionality are implemented, there can be important issues to deal with, such as identifying and visualising potential network inconsistencies.

In addition, we plan to address the protein synonym issue identified by using the Reflect resource [30], a widely-used system that maintains a community-edited dictionary of protein synonyms.

1.4.3 Using 3D Structure Information

In our initial Minardo layout (Fig. 1.1), phosphosites are only represented using protein name and residue numbering. This helps create a clean, sparse layout, but in some cases it can be useful to see more detail about a protein. For example, seeing an image corresponding to a protein's 3D structure – when known – can often give biologists insight into its function [29]. To test the utility of such representations, we calculated images for all proteins in the dataset, using residue colouring to indicate

[3] http://locate.imb.uq.edu.au/

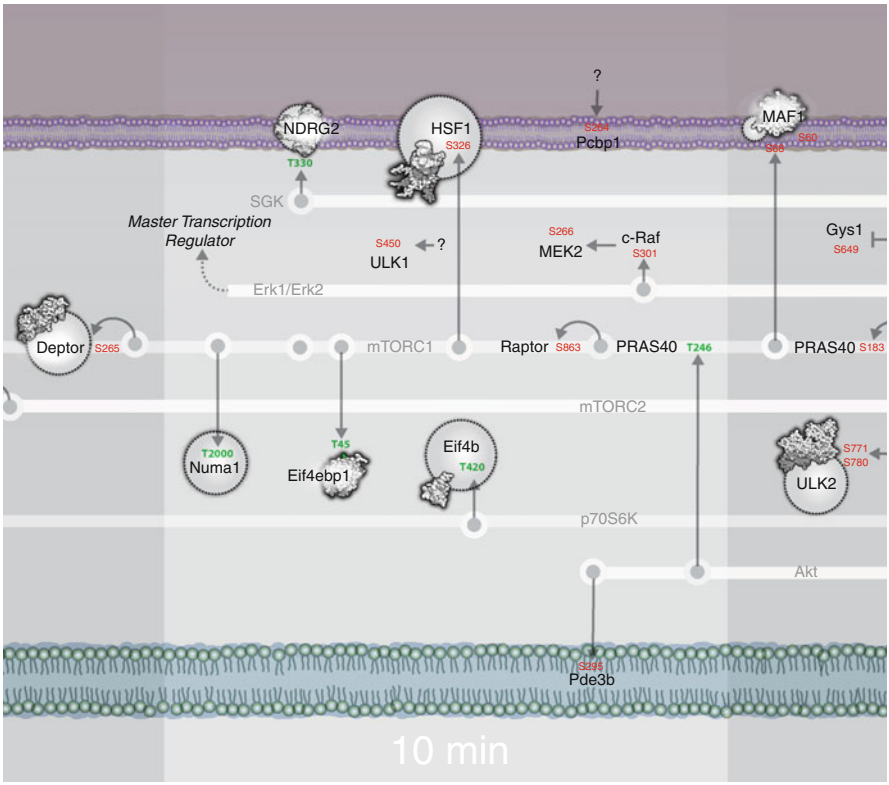


Fig. 1.4 Shown is a segment of Minardo layout with 3D structure images added for proteins. Phospho-residues are coloured using the same scheme as in the heat map, i.e., *red* = serine, *blue* = tyrosine, and *green* = threonine

the location of each phosphosite, then tried adding these images to the network layout. Space filling 3D models were generated using QuteMol [39] from Protein Database (PDB) files, using the PSSH database [34] to find solved structures that are most closely related to our protein of interest. Each 3D model was manually rotated so that the final 2D view shows all phosphosites, if possible. Figure 1.4 shows the results of adding such images to one region of the layout.

Phosphosites are usually solvent accessible, and visualising them on 3D structures can give insight that helps validate them as substrates [40]. Thus, enhancing the visualisation in this way can assist users in a more detailed interpretation of the experimental dataset.

However, we found a key limitation of this approach, which is that many of the phosphosites occur in regions that have no detectable sequence similarity to proteins with known structure – for the complete dataset, this occurs in about half of the 104 sites. This is consistent with the observation that many phosphosites occur in intrinsically disordered region, which are hard to determine structurally [8].

1.4.4 Going Beyond Static Roadmaps

Lastly, to enable network analysis, we also plan to investigate combining our visualisation strategy with more advanced qualitative modelling methods, adding constraints determined from prior biological knowledge and applying machine learning to reverse engineer functional events in the signalling pathway. The problem of inducing an underlying phosphorylation network based on datasets such as those used in this work can be facilitated using intrinsically qualitative formulations, for example, by reducing the number of parameters that need to be determined from the data.

Such formulations allow the use of well-established methods such as Petri Nets to determine properties of states and trajectories, such as the identification of cycles, or unreachable states [12]; in addition, they enable the use of network inference algorithms to reverse engineer hypothesised systems models from data [37].

We are interested in applying such methods to identify from a very large space of phospho-forms those forms most likely to be present in the data, by extending existing techniques that can make use of such constraints [33] to a knowledge-based framework.

1.4.5 Visualisation for Multiple Experiments

As outlined in Sect. 1.3.2 describing our entry in the DREAM8 competition for a proposed workflow to enable comparison of networks under multiple experimental conditions, Minardo's unique features can have several advantages that can be leveraged in such settings. As further work we will investigate how best to implement the proposed extensions to Minardo for the multiple experiment setting.

1.4.6 Limitations

One limitation of the current design is that the order of events shown within one time interval imply a linear ordering in time, which can be too restrictive and, in general, incorrect. In future versions, we plan to investigate strategies to address this limitation, e.g., basing order of placement within an interval on more precise estimates derived from the time-series data.

While the Minardo layout appears to be an improvement on existing solutions to proteomics time-series data, it currently does not address the harder and more general problem of how to integrate data from other experiments, such as measurements of protein and transcript abundance. As a further extension of this work, we plan to test if our approach can accommodate some of the existing approaches to visualising these additional data, such as node colouring or decoration [19].

Conclusions
The Minardo layout provides a novel combination of principles from visual analytics in a customised layout loosely based on cell topology, but strictly ordered first by time, then by causality. This gives life scientists a familiar and helpful frame of reference for organising and interpreting proteomics time-course data. The layout has proven to be useful, leading to new insight into the insulin response pathway. Following the success of our initial research prototype, we are working on generalising the Minardo layout to accommodate similar datasets related to other signalling pathways.

Acknowledgements We gratefully acknowledge helpful conversations with our colleagues Prof. David James, Dr Sean Humphrey, Dr Adelle Coster, Annabel Minard, and Beverley Murrow.

References

1. M. Baitaluk, M. Sedova, A. Ray, A. Gupta, Biological networks: visualization and analysis tool for systems biology. Nucleic Acids Res. **34**, W466–W471 (2006)
2. A. Barsky, J. Gardy, R. Hancock, T. Munzner, Cerebral: a cytoscape plugin for layout of and interaction with biological networks using subcellular localisaiton annotation. Bioinformatics **28**(8), 1040–1042 (2007)
3. V. Batagelj, A. Mrvar, Pajek – program for large network analysis. Connections **2**, 47–57 (1998)
4. E.R. Tufte, *The Visual Display of Quantitative Information*, 2nd edn. (Graphics, Cheshire, 2001)
5. M. Bostock, V. Ogievetsky, J. Heer, D3: Data-driven documents, in *IEEE Transactions Visualization & Computer Graphics (Proceedings of InfoVis 2011)*, Providence, 2011
6. E.G. Cerami, B.E. Gross, E.Demir, I.Rodchenkov, Ö. Babur, N. Anwar, N. Schultz, G.D. Bader, C. Sander, Pathway commons, a web resource for biological pathway data. Nucleic Acids Res. **39**(suppl 1), D685–D690 (2011)
7. P. Chandler, J. Sweller, Cognitive load theory and the format of instruction. Cogn. Instr. **8**(4), 293–332 (1991)
8. T. Chouard, Structural biology: breaking the protein rules. Nature **471**, 151–153 (2011)
9. P. Cohen, The origins of protein phosphorylation. Nat. Cell Biol. **4**, E127–E130 (2002)
10. J. Cox, M. Mann, MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. Nat. Biotechnol. **26**(Database issue), 1367–1372 (2008)
11. M.E. Cusick, N. Klitgord, M. Vidal, D.E. Hill, Interactome: gateway into systems biology. Hum. Mol. Genet. **14**(suppl 2), R171–R181 (2005)
12. R. David, H. Alla, *Discrete, Continuous, and Hybrid Petri Nets*, 2nd edn. (Springer, Berlin, 2010)
13. H. Dinkel, C. Chica, A. Via, C.M. Gould, L.J. Jensen, T.J. Gibson, F. Diella, Phospho.ELM: a database of phosphorylation sites. Nucleic Acids Res. **39**(Database issue), D261–D267 (2011)
14. N. Gehlenborg, S.I. O'Donoghue, N. Baliga, A. Goesmann, M. Hibbs, H. Kitano, O. Kohlbacher, H. Neuweger, R. Schneider, D. Tenenbaum, A-C. Gavin., Visualisation of omics data for systems biology. Nat. Methods **7**, S56–S68 (2010)

15. S. Humphrey, G. Yang, P. Yang, D. Fazakerley, J. Stockli, J. Yang, D. James, Dynamic adipocyte phosphoporteome reveals that Akt directly regulates mTORC2. Cell Metab. **17**, 1–12 (2013)
16. A. Inselberg, The plane with parallel coordinates. Vis. Comput. **1**(4), 69–91 (1985)
17. James lab, http://www.jameslab.com.au, Last visited: 6 June 2013
18. M. Kanehisa, A database for post-genome analysis. Trends Genet. **13**(9), 375–376 (1997)
19. D. Keim, Information visualisation and visual data mining. IEEE Trans. Vis. Comput. Grap. **8**(1), 1077–2626 (2002)
20. H. Kitano, Systems biology: a brief overview. Science **295**(5560), 1662–1664 (2002)
21. D. Ma, C. Stolte, S. Kaur, M. Bain, S.I. O'Donoghue, Visual analytics of phosphorylation time-series data on insulin response, in *International Symposium on Computational Models for Life Sciences*, Sydney, ed. by C. Sun, T. Bednarz, T.D. Pham, P. Vallotton, D. Wang. AIP Conference Proceedings, 2013, pp. 185–196
22. J. Mackinlay, Automating the design of graphical presentations of relational information. ACM Trans. Graph. **5**(2), 110–141 (1986)
23. M. Magrane, The UniProt consortium, UniProt knowledgebase: a hub of integrated protein data. Database (2011). doi:10.1093/database/bar009
24. L. Matthews, G. Gopinath, M. Gillespie, M. Caudy, D. Croft, B. de Bono, P. Garapati, J. Hemish, H. Hermjakob, B. Jassal, A. Kanapin, S. Lewis, S. Mahajan, B. May, E. Schmidt, I. Vastrik, G. Wu, E. Birney, L. Stein, P. D'Eustach, Reactome knowledgebase of human biological pathways and processes. Nucleic Acids Res. **37**, D619–D622 (2009)
25. H. Mi, S. Dong, A. Muruganujan, P. Gaudet, S. Lewis, P.D. Thomas, PANTHER version 7: improved phylogenetic trees, orthologs, and collaboration with the gene ontology consortium. Nucleic Acids Res. **38**, D240–D244 (2010)
26. L. Michaelis, M.L. Menten, Die kinetik der invertinwirkung. Biochemische Zeitschrift **49**, 333–369 (1913)
27. D. Nishimura, A view from the web: biocarta. Biotech Softw. Internet Rep. **2**(3), 117–120 (2001)
28. S.I. O'Donoghue, A.-C. Gavin, N. Gehlenborg, D.S. Goodsell, J.-K. Hériché, C.B. Nielsen, C. North, A.J. Olson, J.B. Procter, D.W. Shattuck, T. Walter, B. Wong, Visualizing biological data – now and in the future. Nat. Methods **7**, S2–S4 (2010)
29. S.I. O'Donoghue, D.S. Goodsell, A.S. Frangakis, F. Jossinet, R.A. Laskowski, M. Nilges, H.R. Saibil, A. Schafferhans, R.C. Wade, E. Westhof, A.J. Olson, Visualization of macromolecular structures. Nat. Methods **7**, S42–S55 (2010)
30. E. Pafilis, S.I. O'Donoghue, L.J. Jensen, H. Horn, M. Kuhn, N.P. Brown, R. Schneider, Reflect: augmented browsing for the life scientist. Nat. Biotechnol. **27**, 308–310 (2009)
31. G.A. Pavlopoulos, S.D. Hooper, A. Sifrim, R. Schneider, J. Aerts, Medusa: a tool for exploring and clustering biological networks. BMC Res. Notes **4**, 104 (2011)
32. S. Pletscher-Frankild, J.X. Binder, K. Tsafou, C. Stolte, S.I. O'Donoghue, R. Schneider, L.J. Jensen, Compartments: unification and visualization of protein subcellular localization evidence. Database (2014). doi:10.1093/database/bau012
33. R. Samaga, J. Saez-Rodriguez, L.G. Alexopoulos, P.K. Sorger, S. Klamt, The logic of EGFR/ErbB signaling: theoretical properties and analysis of high-throughput data. PLoS Comput. Biol. **5**(8), e1000438 (2009)
34. A. Schafferhans, J.E.W. Meyer, S.I. O'Donoghue, The PSSH database of alignments between protein sequences and tertiary structures. Nucleic Acids Res. **31**(5), 494–498 (2003)
35. M. Secrier, R. Schneider, Visualizing time-related data in biology, a review. Brief. Bioinformatics (2013). doi:10.1093/bib/bbt021
36. P. Shannon, A. Markiel, O. Ozier, N. Baliga, J. Wang, D. Ramage, N. Amin, B. Schwikowski, T. Ideker, Cytoscape: a software environment for integrated models of biomolecular interaction networks. Genome Res. **13**, 2498–2504 (2003)

37. A. Srinivasan, M. Bain, Knowledge-guided identification of petri net models of large biological systems, in *Proceedings of the 21st International Conference on Inductive Logic Programming (ILP 2011; Revised Selected Papers)*, Windsor Great Park, ed. by S.H. Muggleton, A. Tamaddoni-Nezhad, F.A. Lisi. Volume 7207 of Lecture Notes in Computer Science (Springer, Berlin, 2012), pp. 317–331
38. J. Sweller, Cognitive load theory, learning difficulty, and instructional design. Learn. Instr. **4**(4), 295–312 (1994)
39. M. Tarini, P. Cignoni, C. Montani, Ambient occlusion and edge cueing to enhance real time molecular visualization. IEEE Trans. Vis. Comput. Graph. **12**(5), S42–S55 (2006)
40. A. Zanzoni, D. Carbajo, F. Diella, P.F. Gheradini, A. Tramontano, M. Helmer-Citterich, A. Via, Phospho3D 2.0: an enhanced database of three-dimensional structures of phosphorylation sites. Nucleic Acids Res. **39**(Database issue), D268–D271 (2011)

Chapter 2
Modeling of Testosterone Regulation by Pulse-Modulated Feedback

Per Mattsson and Alexander Medvedev

Abstract The continuous part of a hybrid (pulse-modulated) model of testosterone (Te) feedback regulation in the human male is extended with infinite-dimensional and nonlinear blocks, to obtain the dynamics that better agree with the hormone concentration profiles observed in clinical data. A linear least-squares based optimization algorithm is developed for the purpose of detecting impulses of gonadotropin-releasing hormone (GnRH) from measured concentration of luteinizing hormone (LH). The estimated impulse parameters are instrumental in evaluating the frequency and amplitude modulation functions parameterizing the pulse-modulated feedback. The proposed approach allows for the identification of all model parameters from the hormone concentrations of Te and LH. Simulation results of the complete estimated closed-loop system exhibiting similar to the clinical data behavior are provided.

Keywords Endocrine systems • Modeling • Impulse detection • Pulsatile feedback • Testosterone regulation • Time delay • Impulsive systems

2.1 Introduction

Hormones are signaling substances of endocrine systems that regulate many aspects in the human body, e.g., metabolism, growth and reproductive processes.

In mathematical modeling of endocrine systems, approaches of different complexity level can be taken, depending on the goal of modeling. A plausible way to obtain a simulation model would be then to derive it from the fundamental principles of biology, biochemistry, and physics. This approach has been used in, e.g., modeling of the glucose-insulin feedback system in type 1 diabetes [23], or in simulating the mechanisms of the human menstrual cycle [28]. However, since endocrine systems usually consist of several glands or cites interacting by means

P. Mattsson (✉) • A. Medvedev
Department of Information Technology, Uppsala University, Box 337,
SE-751 05 Uppsala, Sweden
e-mail: per.mattsson@it.uu.se; alexander.medvedev@it.uu.se

C. Sun et al. (eds.), *Signal and Image Analysis for Biomedical and Life Sciences*,
Advances in Experimental Medicine and Biology 823,
DOI 10.1007/978-3-319-10984-8_2

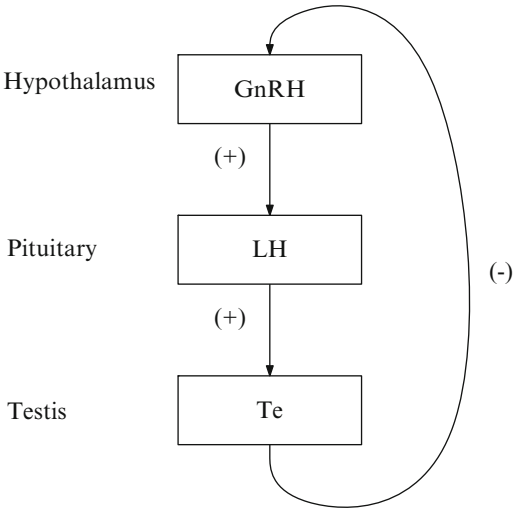


Fig. 2.1 Schematic diagram of the male hypothalamo-piuitary-gonadal system. *Arrows* denote feedforward (stimulatory (+)) and feedback (inhibitory (−)) actions

of hormones, such models are usually of high dimension and very cumbersome to develop. The complexity of the resulting models also makes it hard to populate them with credible parameter values as well as calibrate them to fit biological data.

In order to obtain insights into the principles of biological feedback via mathematical analysis, a modeling approach where only the most essential characteristics and interactions in the system are included appears to be useful. Techniques from system identification [30] can consequently be used to estimate the parameters of the model in a systematic way, yielding a relatively simple model that yet can accurately describe how the hormone concentrations vary with time.

In the endocrine system of testosterone (Te) regulation in the male, schematically depicted in Fig. 2.1, essential roles are played by the luteinizing hormone (LH) and the gonadotropin-realising hormone (GnRH). GnRH is secreted in an episodic and pulsatile manner in the hypothalamus and stimulates the secretion of LH into the blood by the pituitary gland. LH stimulates in turn the secretion of Te in the testes. The endocrine (negative) feedback is then closed by Te inhibiting the secretion of both GnRH and LH [33].

Understanding testosterone regulation is important in, e.g., the treatment of prostate cancer and reproductive failure. There have also been studies showing that the regulatory processes in the GnRH-LH-Te axis are involved in aging and obesity [33].

The structure and function of the pulsatile feedback implemented via GnRH in Te regulation is similar to the action of releasing hormones in other endocrine systems. Typical examples in the human organism that fall into this group are the systems regulating the female sex hormone (estrogen), the stress hormone (cortisol), the growth hormone, insulin, adrenal and parathyroid hormones and some others.

Taking advantage of the relatively simple structure of the GnRH-LH-Te axis, a reductionist approach was taken in, e.g., [29], where a qualitative mathematical

construct describing the male reproductive system was suggested. The model in [29], called the Smith model, can be expressed by the following three, generally nonlinear, ordinary differential equations

$$\begin{aligned} \dot{R} &= f(T) - B_1(R), \\ \dot{L} &= G_1(R) - B_2(L), \\ \dot{T} &= G_2(L) - B_3(T), \end{aligned} \tag{2.1}$$

where $R(t)$, $L(t)$ and $T(t)$ represent the serum concentration of GnRH, LH and Te respectively. The non-negative functions B_1, B_2, B_3 describe the clearing rates of the hormones and f, G_1, G_2 specify the rates of their secretion. Instead of portraying all the complex interactions between the constituting parts of the endocrine system, the Smith model captures only the main features to match the then available mathematical tools.

Naturally, only self-sustained periodical solutions (or chaos) are biologically feasible behaviors of the autonomous model above. The dynamical properties of (2.1) have been studied analytically to great extent for different choices of the functions B_i, G_i, and f. It is relatively common to approximate B_i and G_i by linear functions, so that

$$\begin{aligned} B_i(x) &= b_i x, \quad b_i > 0, \quad i = 1, 2, 3; \\ G_i(x) &= g_i x, \quad g_i > 0, \quad i = 1, 2. \end{aligned} \tag{2.2}$$

In [29], sufficient conditions for (2.1) and (2.2) to have stable periodic solutions are given. When $f(x)$ is chosen to be a Hill function,

$$f(x) = \frac{K}{1 + \beta x^{\rho}}, \tag{2.3}$$

it was shown in [13] that a necessary condition for a periodic solution is that $\rho > 8$. In [27] and [15], it is argued that such a high value of the Hill function order is unrealistic. Indeed, already for $\rho \geq 4$, the Hill function in (2.3) highly resembles a relay characteristic that lacks a proper biological justification.

To resolve the above issue, several attempts have been made to extend the Smith model in such a way that the solutions oscillate for a broader range of the parameters, mostly by introducing special types of nonlinear feedback and time delays. Yet, the Smith model of testosterone regulation in the formulation of [27] is proven to be asymptotically stable for any value of the time delay under a nonlinear feedback in the form of a first-order Hill function [12]. However, by introducing a non-smooth feedback such as piece-wise linear (affine) nonlinearities, multiple periodical orbits and chaos arise in the Smith model [1, 31]. Multiple delays in the Smith model under a second-order Hill function feedback are also shown to lead to sustained nonlinear oscillations in some subspaces of the model parameters [11].

The main adverse property of the classical (differential) Smith model is the equilibrium point of the model dynamics that has to be rendered unstable in order for the system to oscillate. Therefore, in [7], a generalization of the above Smith model was proposed, where the feedback function $f(T)$ is replaced by a pulse modulation operator, to capture the well-known pulsatile nature of GnRH secretion. This version of the Smith model does not have any equilibria. A similar solution with pulse modulation was earlier implemented in the simulation model of the human menstrual cycle [28]. A more general perspective on the pulse-modulated mechanisms in endocrine regulation is provided in [35].

A further improvement of the model in [5] was the introduction of a time delay between LH and Te, to account for the transport phenomena and the time necessary for the secretion of Te in testes. The resulting hybrid and time-delayed model still lends itself to mathematical analysis, and much is already known about its behaviors [6, 26]. It generally exhibits periodical solutions in a wide periodicity range as well as deterministic chaos. The concentration profiles of GnRH and LH produced by the model are consistent with experimental data. For instance, the time-delayed model was successfully used in [16] to estimate GnRH impulses, given measured concentrations of LH. Yet, it seems to be hard to select the closed-loop model parameters in such a way that the simulated Te concentration evolves similarly to the measurements. One reason for this is that, in the model, the secretion of Te never saturates.

In the pulse-modulated Smith model of [7] and [5], the functions B_1, B_2, B_3, G_1 and G_2 were approximated by linear functions, as in (2.2), to facilitate the analysis. Yet, the linear approximations of the secretion functions in (2.1) is, as noted in [3], a gross idealization since, e.g., the saturation of G_2 seems to play an important role in the dynamical behavior of Te, according to the analysis of clinical data in [25]. In [19], a more sophisticated model, where G_1 and G_2 are chosen to be nonlinear saturating functions, is presented. This formulation was also used to extend the model in [5], resulting in a new set of equations that is presented below in Sect. 2.2. Another data-driven and biologically motivated mathematical model for Te regulation has been studied in [19]. Here the saturation of secretion rates is an integral part of the dynamics (Michaelis-Menten kinetics) and the hormone trends fit the measurements well. However, the model is not equipped with a pulse-modulated feedback mechanism and its closed-loop dynamics have not been studied analytically.

When a model structure is selected, the parameters of the model can be determined from measured data by means of system identification algorithms. For this end, techniques from statistics and optimization can be used. In humans, GnRH concentration is typically not measured for ethical reasons. Therefore, the GnRH pulses have to be estimated from measured LH and Te concentrations instead. A state-of-the-art tool for the estimation of hormone profiles in pulsatile endocrine systems is *AutoDecon* described in [18] and freely available on-line. However, this advanced software is tailored to a certain type of open-loop dynamical model and cannot be readily used for other purposes. Therefore, the development of estimation techniques suitable for pulse-modulated (hybrid) models is very much motivated.

In this chapter, the model analyzed in [5] is extended in view of the results reported in [19]. Furthermore, a novel method for estimating GnRH impulses from clinical LH data is introduced and its performance is evaluated. Some initial results on identifying the LH feedforward action on the secretion of Te are also given. Finally, simulation results for the closed-loop system are provided demonstrating good agreement with biological evidence.

2.2 A Pulse-Modulated Mathematical Model of Testosterone Regulation

It is well known [9] that GnRH is released episodically by hypothalamic neurons in modulated secretory bursts. This is not directly reflected by the classical Smith model in (2.1) where the resulting pulsatile temporal profile of the involved hormones is interpreted as smooth nonlinear oscillations. Detailed mathematical descriptions of the GnRH pulses are available, see, e.g., [8]. However, for the purpose of capturing the biologically implemented feedback mechanism, an element implementing pulse-amplitude and pulse-frequency modulation is sufficient, as explained in [35]. Then the problem with the lack of sustained oscillations discussed in Sect. 2.1 can be effectively avoided [7].

Let t_k, $k = 0, 1, 2, \ldots$ be the times when a GnRH pulse is released in the bloodstream, and let λ_k represent the size of the pulse number k. Then a pulse modulated model of testosterone regulation can be written in a state-space form with $x \in \mathbb{R}^3$, and $x_1 = R(t)$, $x_2(t) = L(t)$, $x_3(t) = T(t)$, as

$$\dot{x}(t) = \begin{bmatrix} -b_1 & 0 & 0 \\ g_1 & -b_2 & 0 \\ 0 & g_2 & -b_3 \end{bmatrix} x(t), \quad \text{if } t \neq t_k \tag{2.4}$$

$$x(t_k^+) = x(t_k^-) + \lambda_k \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \qquad \text{if } t = t_k \tag{2.5}$$

where b_1, b_2, b_3, g_1 and g_2 are the positive parameters as in (2.2). The firing times t_k and the impulse weights λ_k are then given by the recursion

$$t_{k+1} = t_k + \tau_k, \quad \tau_k = \Phi(x_3(t)), \quad \lambda_k = F(x_3(t)), \tag{2.6}$$

where $\Phi(\cdot)$ is a frequency modulation characteristic and $F(\cdot)$ is an amplitude modulation characteristic.

The model in (2.4)–(2.6) is referred to as the pulse-modulated Smith model, and was analyzed in detail in [7]. It constitutes a hybrid systems that evolves in continuous time according to (2.4), but undergoes momentous jumps in the continuous state vector at discrete points governed by (2.5). The latter phenomenon corresponds to the secretion of a GnRH impulse.

The dynamics of (2.4)–(2.6) are thoroughly studied and known to exhibit oscillating solutions that are either periodic or chaotic [36]. However, as discussed in [19], the LH feedforward on Te should be exerted via a time average of the LH concentration, and is expected to saturate for high concentrations of LH. Also, the basal secretion of the hormones should be added to the model, as well as the time it takes for LH to travel from the pituitary to the testes should be taken into account. In order to describe this mathematically, let $H_L(t)$ be a time average of LH, i.e.,

$$H_L(t) = \frac{1}{\ell}\int_{t-\ell}^{t} x_2(s)\,\mathrm{d}s, \tag{2.7}$$

where ℓ is the length of the sliding window on which the average value of x_2 is calculated. Also let $Z_T(t)$ be the secretion rate of Te at time t. Then

$$Z_T(t) = g_2\Phi_L(H_L(t-\tau)), \tag{2.8}$$

where $\Phi_L(\cdot)$ is a saturating function, and τ represents the time it takes for LH to travel to the testes.

Now the model can then be summarized as

$$\dot{x}(t) = A_0x(t) + A_1\Phi_L(H_L(t-\tau)) + \beta, \quad \text{if } t \neq t_k \tag{2.9}$$

$$x(t_k^+) = x(t_k^-) + \lambda_k B, \qquad \text{if } t = t_k \tag{2.10}$$

where

$$A_0 = \begin{bmatrix} -b_1 & 0 & 0 \\ g_1 & -b_2 & 0 \\ 0 & 0 & -b_3 \end{bmatrix}, \quad A_1 = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & g_2 & 0 \end{bmatrix}, \quad B = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}$$

and the vector $\beta \in \mathbb{R}^3$ describes the basal secretion of the hormones. Therefore, the basal secretion is seen as a constant signal exogenous to the endocrine regulation loop. The impulse times t_k and weights λ_k are still given by (2.6). This model was first introduced in [25].

The frequency and amplitude modulation functions $\Phi(\cdot)$, and $F(\cdot)$, as well as the saturating function $\Phi_L(\cdot)$, are here assumed to be Hill functions of the form

$$\begin{aligned} \Phi(y) &= k_1 + k_2\frac{(y/h)^p}{1+(y/h)^p}, \\ F(y) &= k_3 + \frac{k_4}{1+(y/h)^p}, \\ \Phi_L(H_L) &= \frac{(H_L/h_L)^{p_L}}{1+(H_L/h_L)^{p_L}}, \end{aligned} \tag{2.11}$$

where $p, p_L \in \mathbb{N}^+$ are the (positive integer) orders of the Hill functions and $k_1, k_2, k_3, k_4, h, h_L \in \mathbb{R}^+$. Notice that all three functions in (2.11) are monotone and bounded.

2.3 Parameter Estimation

Given the model defined in Sect. 2.2, it is of interest to estimate the unknown parameters using experimental data. This is usually called model calibration and is carried out by algorithms developed in system identification [30].

In the case of the GnRH-LH-Te axis in the human male, the data usually consist of the LH and Te concentrations measured at certain instants over time. The concentration of LH and Te can be assessed from blood samples and are often measured every 10 min or similar. GnRH is usually not measured in the human, since it is not possible to do so in a non-invasive way, see, e.g., [22].

The unknown parameters in the model expressed by (2.9)–(2.10) are the constants b_1, b_2, b_3, g_1 and g_2 in the matrices A_0 and A_1. The nonlinear functions in (2.11) also contain unknown quantities.

To the best of knowledge of these authors, there presently exists no method for identifying the hybrid feedback system in (2.9)–(2.10). For this reason, only the open-loop system dynamics are considered here and the pulse modulation functions are not explicitly identified from data. Nevertheless, estimates of the unknown parameters in (2.11), as will be shown further, can be readily obtained from the estimates of the weights λ_k and the firing times t_k of the GnRH pulses.

2.3.1 Estimating the GnRH Impulses

Since the GnRH concentration is in practice an unobserved signal in (2.9), it is of interest to estimate the GnRH impulse times and weights from measured LH data. The half-life times b_1 and b_2 of GnRH and LH, respectively, are also considered to be unknown here, so these parameters have to be estimated in some way. Naturally, it is known from biochemistry to what range of values these quantities belong in vitro but they are also known to be subject to significant inter-individual variation in vivo.

Before going into details about the estimation procedure, notice first that it is enough to study the first two states of x in (2.9) when the effect of an GnRH impulse on LH is considered. Thus, introduce the reduced (continuous) state vector $\tilde{x}(t) = \begin{bmatrix} x_1(t) & x_2(t) \end{bmatrix}^T$. Also notice that the jump in the state-vector at time t_k, as in (2.10), equivalently can be represented as an input to the system in the form of a delayed Dirac delta impulse, $\delta(t - t_k)$. Therefore, it follows from (2.9) to (2.10) that

$$\dot{\tilde{x}} = \tilde{A}\tilde{x} + \tilde{B}\xi(t) + \tilde{\beta}, \tag{2.12}$$

$$x_2(t) = \tilde{C}\tilde{x}(t), \tag{2.13}$$

where

$$\tilde{A} = \begin{bmatrix} -b_1 & 0 \\ g_1 & -b_2 \end{bmatrix}, \quad \tilde{B} = \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \quad \tilde{\beta} = \begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix}, \quad \tilde{C} = \begin{bmatrix} 0 & 1 \end{bmatrix}, \tag{2.14}$$

and

$$\xi(t) = \sum_{k=0}^{\infty} \lambda_n \delta(t - t_k). \tag{2.15}$$

Equations (2.12) and (2.13) can be seen as a linear time invariant (LTI) system with the input vector $\tilde{B}\xi(t) + \beta$. If (2.12) is subject to the initial condition $\tilde{x}(t_o) = \tilde{x}_o = \begin{bmatrix} x_{o,1} & x_{o,2} \end{bmatrix}^T$, then the solution to the differential equations in (2.12) is [14]

$$x_2(t) = \tilde{C}\left(\mathrm{e}^{\tilde{A}(t-t_o)}\tilde{x}_o + \int_{t_o}^{t} \mathrm{e}^{\tilde{A}(t-s)}(\tilde{\beta} + \tilde{B}\xi(s))\mathrm{d}s \right).$$

By evaluating the integral, the expression above can be rewritten as

$$x_2(t) = \tilde{C}\mathrm{e}^{\tilde{A}(t-t_o)}\tilde{x}_o + \tilde{C}\tilde{A}^{-1}\left(\mathrm{e}^{\tilde{A}(t-t_o)} - I\right)\tilde{\beta} + \sum_{k=0}^{\infty} g_1 \lambda_k z(t - t_k), \tag{2.16}$$

where

$$z(t) = \frac{\mathrm{e}^{-b_2 t} - \mathrm{e}^{-b_1 t}}{b_1 - b_2} H(t),$$

and $H(\cdot)$ is the Heaviside step function.

From (2.16), it can be seen that the unknown parameter g_1 always appears in a product with other unknown parameters, e.g., the impulse weights λ_k [16]. For this reason it is not possible to uniquely determine g_1 from measured data. Similarly, it is in practice not possible to separate β_1 from β_2 when the data are collected from the closed-loop system. This makes sense intuitively: Since the basal secretion rate of GnRH is unknown, it is not possible to distinguish between the basal secretion of LH and the LH secreted due to basal GnRH secretion. For these reasons it will be assumed that $g_1 = 1$ and $\beta_1 = 0$. Thus, any estimated GnRH signal should be considered as a virtual signal, providing information about the GnRH firing times t_k and the ratios between the weights λ_k. However, the actual concentrations might differ in the real system.

2.3.1.1 Estimating Firing Times and Weights

A method for identifying the GnRH pulse parameters in (2.12) was proposed in [17]. However, the method, in its present form, is restricted to the identification of two

impulses at a time. A general method for estimating unknown input impulses was also developed in [24], but it assumes continuous access to the output data, and is therefore hard to apply directly to LH data where the sampling is typically slow (e.g., 10 min).

In this section a method that can estimate GnRH impulses from longer and undersampled datasets, originally introduced in [25], is described. In this method the specific structure of (2.12)–(2.13) is directly used to recover information on the GnRH pulses.

Now assume that the elimination rates of GnRH and LH, i.e., b_1 and b_2, are known. Exactly how these can be found is discussed in Sect. 2.3.1.2. Also assume that $b_1 > b_2$, i.e., that the elimination of GnRH is faster than the elimination of LH. This is clearly the case in the GnRH-LH-Te axis, see, e.g., [19].

Now let $\hat{t}_0, \hat{t}_1, \ldots, \hat{t}_m$ be the times when the LH concentration is measured, i.e., the instants when the blood samples are taken. Also let $t_1, \ldots, t_n$ be the all the impulse times in the interval $(0, \hat{t}_m)$, i.e., $t_1, \ldots, t_n$ denotes all times when a GnRH impulse is released from the hypothalamus during the measurement period.

The goal is then to recover the impulse times t_k and impulse weights λ_k, $k = 1, \ldots, n$, from the measured LH concentrations $x_2(\hat{t}_l)$, $l = 1, \ldots, m$.

A common way to estimate unknown parameters is to find parameters that minimize the square of the error between measured data and the output produced by the model. The problem with doing so in the present case is that the number of impulses, as well as the impulse times, are unknown. This makes the problem non-convex, so that no general method for finding a global minimum exists.

However, if it is, to start with, assumed that impulse times are known then it turns out that the impulse weights λ_k can be estimated by solving a nonnegative linear least-squares optimization problem. To see this, let

$$z_i = \left[(1 - \mathrm{e}^{-b_2\hat{t}_i})/b_2 \; e^{-b_2\hat{t}_i} \; z(\hat{t}_i) \; z(\hat{t}_i - t_1) \cdots z(\hat{t}_i - t_n)\right], \quad Z = \left[z_1^T \cdots z_m^T\right]^T,$$

$$G = \left[\beta_2 \; x_{o,2} \; x_{o,1} \; \lambda_1 \cdots \lambda_n\right]^T, \quad \text{and} \quad Y = \left[x_2(\hat{t}_0) \cdots x_2(\hat{t}_m)\right].$$

Recalling from Sect. 2.3.1 the assumptions that $g_1 = 1$ and $\beta_1 = 0$, it follows that

$$\tilde{C}e^{\tilde{A}(t-t_o)}\tilde{x}_o = x_{o,1}z(t - t_o) + x_{o,2}e^{-b_2(t-t_o)},$$
$$\tilde{C}\tilde{A}^{-1}\left(e^{\tilde{A}(t-t_o)} - I\right)\tilde{\beta} = \left(1 - e^{-b_2(t-t_o)}\right)\frac{\beta_2}{b_2}.$$

With the data vector Y given by (2.16), it is now straightforward to validate that

$$ZG = Y.$$

Thus, the impulse weights, initial conditions and basal secretion levels can all be found by solving a system of linear algebraic equations. However, as Y consists of measured data, it usually corrupted by disturbances due to measurement noise etc.,

so that the system of linear equations above might not have a solution. In this case, the following optimization problem has to be solved instead

$$\begin{aligned} \min_{G} \quad & \|ZG - Y\|_2^2, \\ \text{s.t.} \quad & G \geq 0, \end{aligned} \tag{2.17}$$

where the inequality in the constraint is understood element-wise and follows from the biology of the system. This is a nonnegative least-squares optimization problem and can be solved efficiently with standard methods, see, e.g., [2].

The problem now is that the impulse times t_k are unknown. A practical technique that can then be utilized is to create a grid in the interval $(\hat{t}_0, \hat{t}_m)$, and assume that the impulse times belong to this grid. Provided there are grid points close enough to the true impulse times, the solution to (2.17) will hopefully assign positive impulse weights to the grid points close to the true impulse times and zero impulse weights to the other grid points. The question is still, how should the grid points be chosen? One approach that works out well in the case studied here is to follow the following algorithm.

Algorithm 1
1. Solve (2.17) with all grid points equal to the sampling times.
2. If two consecutive grid points $\hat{t}_l$ and $\hat{t}_{l+1}$ have been assigned the non-zero weights $\hat{\lambda}_l$ and $\hat{\lambda}_{l+1}$, replace those impulses with one impulse fired at the time

$$t_k = \hat{t}_l + \frac{1}{b_2 - b_1} \ln\left(1 + \frac{\hat{\lambda}_{l+1}\left(e^{b_2\Delta} - e^{b_1\Delta}\right)}{\hat{\lambda}_l + \hat{\lambda}_{l+1}e^{b_1\Delta}}\right)$$

and with the impulse weight

$$\lambda_k = e^{-b_1(t_k - \hat{t}_l)}\left(\hat{\lambda}_{l+1}e^{b_1\Delta} + \hat{\lambda}_l\right),$$

where $\Delta = \hat{t}_{l+1} - \hat{t}_l$.

To motivate Algorithm 1, assume first that the data vector Y is generated from (2.16) exactly and contains no disturbances. In the algorithm, the grid points are chosen equal to the sampling times and (2.17) is used to find an optimal set of impulse weights. It is straightforward to show that when the data vector Y is generated from (2.16) without disturbances, then, with this set of grid points, the optimal solution to (2.17) results in the cost function equal to zero.

Finally, if two consecutive grid points received non-zero impulse weights, those impulses are combined into a single impulse in between the two grid points. It can be shown that this merging of impulses will not change the solution of (2.16) at the sampling times, even if the solution changes between them. Hence the resulting error (and the cost function) will still be the same. In the case of no disturbances, it thus follows that the cost function after the merging of impulses is still zero. If it

additionally assumed that there are at least three sampling times in between any two impulses in (2.16), then there will be no ambiguity about how the impulses should be merged. From the above discussion, the following proposition can be proven.

Proposition 1 *Assume that the elements of Y are generated according to* (2.16)*, it holds that $b_1 > b_2 > 0$, and that there are at least three sampling instances in between any two firing times. Then Algorithm 1 yields the true firing times and impulse weights.*

Notice here that the assumption about three sampling times in between any two firing times is only needed to guarantee the uniqueness of the solution. Algorithm 1 is feasible even when the firing times are closer to each other and it will produce an optimal solution in the least-squares sense.

In real data, however, the measurements comprising Y will be corrupted by disturbances. In that case, following Algorithm 1 will typically result in solutions where every grid point is assigned a non-zero impulse weight. If the data are, for instance, sampled every 10 min, this is not biologically reasonable since there is usually more than 1 h between two GnRH impulses in the true system, see, e.g., [20]. The reason why the algorithm behaves this way is that the cost function can be improved by fitting, e.g., the measurement noise by adding false impulses.

A way to alleviate this problem is to constrain the solutions to (2.17) in such a way that at most $r \in \mathbb{N}^+$ grid points are assigned a non-zero impulse weight. That is, the solution of the optimization problem is restricted to a sparse set of non-zero impulses weights. However, the resulting optimization problem is in general very hard to solve. Instead, the sparsity constraint can be efficiently enforced by putting a constraint on the sum of the (positive) impulse weights, i.e., the l_1-norm, see [32]. This results in the following optimization problem

$$\begin{aligned} \min_{G} \quad & \|ZG - Y\|_2^2, \\ \text{s.t.} \quad & \sum \lambda_k < \lambda_{\max}, \\ & G \geq 0, \end{aligned} \tag{2.18}$$

that can be solved efficiently by existing methods, see [2]. The optimal solutions of (2.18) are not in principle guaranteed to be sparse, but they often are. In the case studied here, i.e., estimating GnRH impulses from LH data, it is usually possible to find a value of $\lambda_{\max}$ that gives sparse solutions with a good fit, as will be seen in the forthcoming sections.

2.3.1.2 Estimating the Parameters

In order to find the GnRH impulses as in Sect. 2.3.1.1, the parameters b_1 and b_2 in (2.14) are needed. That is, the linear system in (2.12)–(2.13) has to be identified. If both the input $\xi(t)$ and the output $x_2(t)$ were known, then methods from system

identification of linear systems would be directly applicable, see [30]. However, this is not the case here since the impulse times, and thus the input function $\xi(t)$, are unknown.

According to [19], the parameters b_1 and b_2 should satisfy the bounds

$$\begin{aligned} 0.23\ \text{min}^{-1} \leq b_1 \leq 0.69\ \text{min}^{-1}, \\ 0.0087\ \text{min}^{-1} \leq b_2 \leq 0.014\ \text{min}^{-1}. \end{aligned} \tag{2.19}$$

A simple method to pick an estimate of these parameters is thus to chose a fine grid over the intervals in (2.19). Then, for each combination of b_1 and b_2 on these grids, a set of impulses is computed as in Sect. 2.3.1.1. Finally the parameter pair that gives the smallest squared error is picked as an estimate of b_1 and b_2. This is the method used in [25], and it is also used in Sect. 2.4 to estimate the GnRH impulses from clinical LH data. An alternative to this very practical approach is a Laguerre domain identification technique suggested in [17].

2.3.2 Estimating the Testosterone Dynamics

Finally, the dynamics from LH to Te have to be identified. From (2.9) to (2.10), it follows that

$$\dot{T}(t) = \dot{x}_3(t) = -b_3 x_3(t) + g_2 \Phi_L(H_L(t-\tau)) + \beta_3. \tag{2.20}$$

This system can thus be seen as a Hammerstein system with the input signal $H_L(t-\tau)$, i.e., a system that consists of a static nonlinearity followed by a linear system. A common method for identification of Hammerstein systems is to introduce overparameterization and transform the model into a linear multiple-input-single-output system [4]. However, this method introduces several new parameters and can lead to highly uncertain estimates for small datasets. Methods for recursive Hammerstein identification that do not resort to model overparameterization are presented in [10].

However, for the purpose of illustration, the method used here is to simply grid over a domain of the unknown parameters h_L, p_L, τ and ℓ in the nonlinear part of (2.20). For each combination of nonlinear parameters, the parameters b_3, b_2 and β_3 as well as the initial conditions, are determined with the ordinary least-squares method, see [30].

2.4 Experimental Results

The methods developed in this chapter have been tested on LH and Te concentrations measured in 18 healthy human males. The data were collected for 17 h and sampled every 10 min, see [21] for a description of the data and experimental protocol.

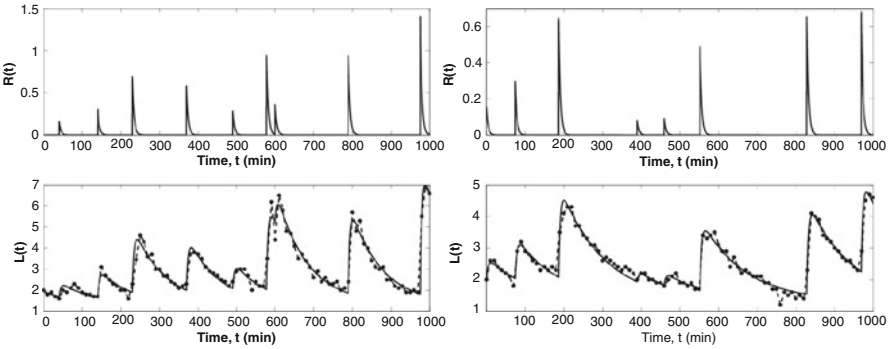


Fig. 2.2 Result of GnRH impulses estimation. *Upper row* shows the estimated virtual GnRH levels for each patient. *Bottom row* shows the estimated LH levels (*solid line*), together with the measured data (*dashed line* with *dots* at each sampling time). The data in the *left column* are from a patient that is 27 years old, and in the *right column* from a 40 years old patient (Reprinted with permission from [25]. Copyright 2013, AIP Publishing LLC)

For all patients, the LH data were well explained by the model in (2.12), with a sparse set of 4-10 impulses in 1,000 min which appears to be biologically feasible. For illustration, the results of two patients are shown in Fig. 2.2.

To test the identification of the testosterone dynamics, the mathematical model in (2.20) was simulated with the measured LH as input. It should be noted that such a simulation cannot be considered a complete estimation of the testosterone concentration since it only captures the Te secreted due to the LH feedforward signal and the constant basal secretion. In the actual endocrine system, the concentration of Te is also involved in other regulations and events outside the GnRH-LH-Te loop, both of endocrine and non-endocrine nature [37]. However, the simulation results with measured LH data indicate that the simulated Te concentration still follows the general trend in the corresponding measured Te data. For illustration, the result for the same two patients as above are shown in Fig. 2.3. Yet, the identification of Te still needs to be studied in more detail, both when it comes to the model structure and the identification methods.

2.5 Simulations of the Closed-Loop System

In this section, the complete model described by (2.9)–(2.11) is simulated. In fact, the authors have failed to find in the literature any other identified from data model of the GnRH-LH-Te axis that is tested in closed loop. The model parameters are estimated from of the LH and Te data collected from the same 27 years old and a 40 years old healthy individuals as mentioned before. The modulation functions in (2.6) were then chosen so that the minimum and maximum time between impulses corresponded to the minimum and maximum inter-impulse times in the estimated GnRH data.

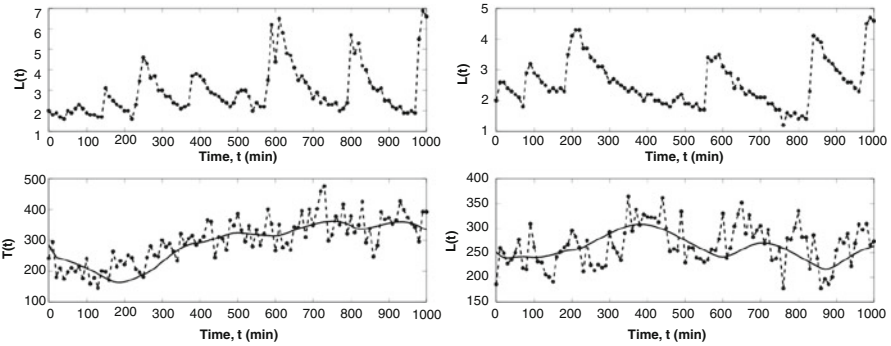


Fig. 2.3 Result of Te dynamics identification. The *upper row* shows the measured LH data that are used as input. The *bottom row* shows measured Te (*dashed line*) together with Te concentration given by simulation of the model (*solid line*). The *left column* corresponds to data from a 27 years old patient, and the *right column* from a 40 years old patient (Reprinted with permission from [25]. Copyright 2013, AIP Publishing LLC)

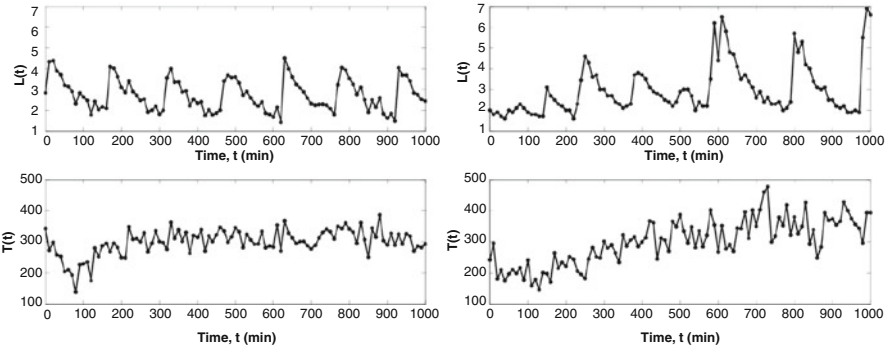


Fig. 2.4 Simulation of the closed-loop system. *Left column*: simulation; *right column*: measured data in a 27 years old patient (Reprinted with permission from [25]. Copyright 2013, AIP Publishing LLC)

To imitate the clinical conditions, the simulated data were sampled every 10 min with white Gaussian measurement noise added. Figures 2.4 and 2.5 compare the results of the closed-loop simulation with the experimental data that the parameters were estimated from. Since the modulation functions in the feedback are not formally identified from measurements, it cannot be expected that the simulated data and the real data should look exactly the same. However, it can be seen that the simulated and real data in many ways exhibit similarity. For example, the numbers of impulses are about the same, the hormone concentration are of the same level, etc.

One important difference is that the amplitude of the LH pulses seem to vary more over time in the real data which phenomenon can be attributed to a prominent circadian rhythm in Te concentrations [34]. Circadian rhythm can be incorporated by several means in the model developed in this chapter. For example, the amplitude

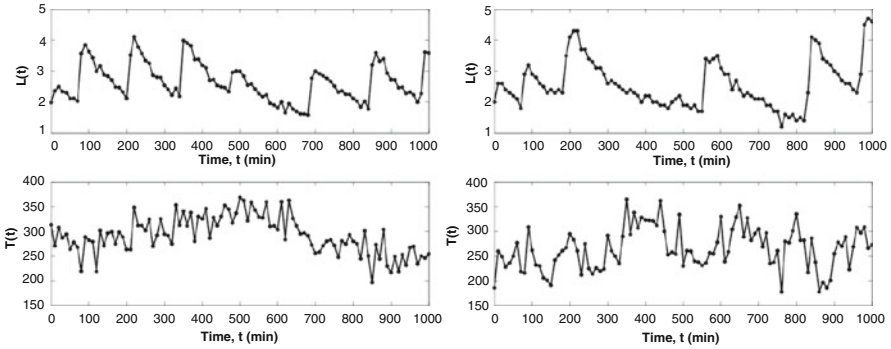


Fig. 2.5 Simulation of the closed-loop system. *Left column*: simulation; *right column*: measured data in a 40 years old patient (Reprinted with permission from [25]. Copyright 2013, AIP Publishing LLC)

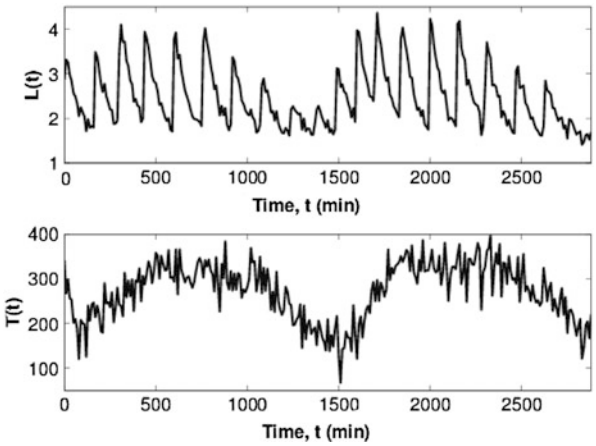


Fig. 2.6 Simulation over 48 h, with simulated circadian rhythm (Reprinted with permission from [25]. Copyright 2013, AIP Publishing LLC)

modulation function could depend on time in a periodic manner. However, to study such slow periodic behavior, it would be desirable to investigate datasets that span over at least two full days. Figure 2.6 presents a simulation example with the same parameters as in Fig. 2.4, except that the amplitude modulation functions is multiplied by a sinusoid function with a period of 24 h.

Conclusions and Future Work
An extension of a hybrid mathematical model of testosterone regulation is suggested, together with an algorithm for estimating the firing times of GnRH impulses from time series of measured LH concentrations.

(continued)

Tests on hormone data collected in healthy human males indicate that the model explains the experimental LH concentration profile very well with a sparse set of GnRH impulses. An initial study on system identification of the testosterone dynamics is also presented. It is demonstrated that the model captures the excursions in the Te secretion caused by the alternations of LH concentration. However, further research into the Te secretion model structure, along with more refined identification methods, are still needed.

With the parameter estimates obtained by identification from experimental data, the simulated complete closed-loop system exhibits behaviors similar to clinical data. Yet formal methods to identify the modulation functions in the feedback from Te to GnRH are still needed. Also, the circadian rhythm should be incorporated into the modulation functions to enhance model fidelity. To achieve these goals, new clinical experiments have to be performed.

Acknowledgements This work was in part financed by the European Research Council, Advanced Grant 247035 (SysTEAM) and Grant 2012-3153 from the Swedish Research Council.

References

1. R.H. Abraham, H. Kocak, W.R. Smith, Chaos and intermittency in an endocrine system model, in *Chaos, Fractals, and Dynamics*, ed. by P. Fischer, W.R. Smith, vol. 98 (Marcel Dekker, New York , 1985), pp. 33–70
2. S. Boyd, L. Vandenberghe, *Convex Optimization* (Cambridge University Press, New York, 2004)
3. M. Cartwright, M. Husain, A model for the control of testosterone secretion. J. Theor. Biol. **123**(2), 239–250 (1986)
4. F. Chang, R. Luus, A noniterative method for identification using Hammerstein model. IEEE Trans. Autom. Contr. **16**(5), 464–468 (1971)
5. A. Churilov, A. Medvedev, P. Mattsson, Analysis of a pulse-modulated model of endocrine regulation with time-delay, in *IEEE 51st Annual Conference on Decision and Control (CDC)*, Maui (2012), pp. 362–367
6. A. Churilov, A. Medvedev, P. Mattsson, On finite-dimensional reducibility of time-delay systems under pulse-modulated feedback, in *IEEE 52nd Annual Conference on Decision and Control (CDC)*, Firenze (2013), pp. 362–367
7. A. Churilov, A. Medvedev, A. Shepeljavyi, Mathematical model of non-basal testosterone regulation in the male by pulse modulated feedback. Automatica **45**(1), 78–85 (2009)
8. F. Clément, J.-P. Françoise, Mathematical modeling of the GnRH-pulse and surge generator. SIAM J. Appl. Dyn. Syst. **6**(2), 441–456 (2007)
9. D.J. Dierschke, A.N. Bhattacharya, L.E. Atkinson, E. Knobil, Circhoral oscillations of plasma LH levels in the ovariectomized rhesus monkey. Endocrinology **87**, 850–853 (1970)
10. F. Ding, X.P. Liu, G. Liu, Identification methods for Hammerstein nonlinear systems. Digit. Signal Process. **21**(2), 215–238 (2011)
11. D. Efimov, A. Fradkov, Oscillatority conditions for nonlinear systems with delay. J. Appl. Math. (2007). article ID 72561

12. G. Enciso, E.D. Sontag, On the stability of a model of testosterone dynamics. J. Math. Biol. **49**, 627–634 (2004)
13. L.S. Farhy, Modeling of oscillations in endocrine networks with feedback, in *Numerical Computer Methods, Part E*, ed. by M.L. Johnson, L. Brand. Methods in Enzymology, vol. 384 (Academic, San Diego,2004), pp. 54–81
14. T. Glad, L. Ljung, *Control Theory: Multivariable and Nonlinear Methods* (Taylor and Francis, New York/London, 2000)
15. W.J. Heuett, H. Qian, A stochastic model of oscillatory blood testosterone levels. Bull. Math. Biol. **68**(6), 1383–1399 (2006)
16. E. Hidayat, A. Medvedev, Identification of a pulsatile endocrine model from hormone concentration data, in *IEEE International Conference on Control Applications (CCA)*, Dubrovnik (2012), pp. 356–363
17. E. Hidayat, A. Medvedev, Laguerre domain identification of continuous linear time-delay systems from impulse response data. Automatica **48**(11), 2902–2907 (2012)
18. M.L. Johnson, L. Pipes, P.P. Veldhuis, L.S. Farhy, R. Nass, M.O. Thorner, W.S. Evans, AutoDecon: a robust numerical method for the quantification of pulsatile events, in *Computer Methods, Volume A*, ed. by M.L. Johnson, L. Brand. Methods in Enzymology, vol. 454 (Elsevier, Amsterdam, 2009), pp. 367–404
19. D.M. Keenan, J.D. Veldhuis, A biomathematical model of time-delayed feedback in the human male hypothalamic-pituitary-leydig cell axis. Am. J. Physiol. Endocrinol. Metab. **275**(1), E157–E176 (1998)
20. D.M. Keenan, S. Chattopadhyay, J.D. Veldhuis, Composite model of time-varying appearance and disappearance of neurohormone pulse signals in blood. J. Theor. Biol. **236**(3), 242–255 (2005)
21. D.M. Keenan, I. Clarke, J.D. Veldhuis, Non-invasive analytical estimation of endogenous gonadotropin-releasing hormone (GnRH) drive: analysis using graded competitive GnRH-receptor antagonism and a calibrating pulse of exogenous GnRH. Endocrinology **152**(12), 4882–4893 (2011)
22. D.M. Keenan, W. Sun, J.D. Veldhuis, A stochastic biomathematical model of male reproductive hormone systems. SIAM J. Appl. Math. **61**(3), 934–965 (2000)
23. B.P. Kovatchev, M. Breton, C. Dalla Man, C. Cobelli, In silico preclinical trials: a proof of concept in closed-loop control of type 1 diabetes. J. Diabetes Sci. Technol. **3**(1), 44–55 (2009)
24. P. Mattsson, A. Medvedev, Estimation of input impulses by means of continuous finite memory observers, in *American Control Conference*, Montreal, June 2012, pp. 6769–6774
25. P. Mattsson, A. Medvedev, Modeling of testosterone regulation by pulse-modulated feedback: an experimental data study, in *International Symposium on Computational Models for Life Sciences*, ed. by C. Sun, T. Bednarz, T.D. Pham, P. Vallotton, D. Wang. AIP Conference Proceedings, Sydney, vol. 1559 (2013), pp. 333–342
26. A. Medvedev, A. Churilov, P. Mattsson, Periodical solutions in a pulse-modulated model of endocrine regulation with time-delay. IEEE Trans. Autom. Contr. **59**(3), 728–733 (2014)
27. J.D. Murray, *Mathematical Biology, I: An Introduction*, 3rd edn. (Springer, New York, 2002)
28. N.L. Rasgon, L. Pumphrey, P. Prolo, S. Elman, A.B. Negrao, J. Licinio, A. Garfinkel, Emergent oscillations in mathematical model of the human menstrual cycle. CNS Spectr. **8**(11), 805–814 (2003)
29. W.R. Smith, Hypothalamic regulation of pituitary secretion of luteinizing hormone. II. Feedback control of gonadotropin secretion. Bull. Math. Biol. **42**(1), 57–78 (1980)
30. T. Söderström, P. Stoica, *System Identification* (Prentice-Hall, Upper Saddle River, 1988)
31. C.T. Sparrow, Chaos in a three-dimensional single loop feedback system with a piecewise linear feedback function. J. Math. Anal. Appl. **83**(1), 275–291 (1981)
32. R. Tibshirani, Regression shrinkage and selection via the LASSO. J. R. Stat. Soc. Ser. B (Methodol.) **58**(1), 267–288 (1996)
33. J.D. Veldhuis, Recent insights into neuroendocrine mechanisms of aging of the human male hypothalamic-pituitary-gonadal axis. J. Androl. **20**(1), 1–18 (1999)

34. J.D. Veldhuis, J.C. King, R.J. Urban, A.D. Rogol, W.S. Evans, L.A. Kolp, M.L. Johnson, Operating characteristics of the male hypothalamo-pituitary-gonadal axis: pulsatile release of testosterone and follicle-stimulating hormone and their temporal coupling with luteinizing hormone. J. Clin. Endocrinol. Metab. **65**(5), 929–41 (1987)
35. J.J. Walker, J.R. Terry, K. Tsaneva-Atanasova, S.P. Armstrong, C.A. McArdle, S.L. Lightman, Encoding and decoding mechanisms of pulsatile hormone secretion. J. Neuroendocr. **22**, 1226–1238 (2009)
36. Z.T. Zhusubaliyev, A.N. Churilov, A. Medvedev, Bifurcation phenomena in an impulsive model of non-basal testosterone regulation. Chaos Interdiscip. J. Nonlinear Sci. **22**(1), 013121 (2012)
37. M. Zitzmann, E. Nieschlag, Testosterone levels in healthy men and the relation to behavioural and physical characteristics: facts and constructs. Eur. J. Endocr. **144**(3), 183–197 (2001)

Chapter 3
Hybrid Algorithms for Multiple Change-Point Detection in Biological Sequences

Madawa Priyadarshana, Tatiana Polushina, and Georgy Sofronov

Abstract Array comparative genomic hybridization (aCGH) is one of the techniques that can be used to detect copy number variations in DNA sequences in high resolution. It has been identified that abrupt changes in the human genome play a vital role in the progression and development of many complex diseases. In this study we propose two distinct hybrid algorithms that combine efficient sequential change-point detection procedures (the Shiryaev-Roberts procedure and the cumulative sum control chart (CUSUM) procedure) with the Cross-Entropy method, which is an evolutionary stochastic optimization technique to estimate both the number of change-points and their corresponding locations in aCGH data. The proposed hybrid algorithms are applied to both artificially generated data and real aCGH experimental data to illustrate their usefulness. Our results show that the proposed methodologies are effective in detecting multiple change-points in biological sequences of continuous measurements.

Keywords Cross-entropy method • Change-point modelling • aCGH data • DNA sequences • Copy number variation • Sequential change-point analysis • Shiryaev-Roberts procedure • Cumulative sum procedure • Combinatorial optimization • Stochastic optimization

---

This work was carried out when the author was at the Department of Mathematics, Mari State University, Russia

M. Priyadarshana (✉) • G. Sofronov
Faculty of Science, Department of Statistics, Macquarie University, Sydney, NSW 2109, Australia
e-mail: madawa.weerasinghe@mq.edu.au; georgy.sofronov@mq.edu.au

T. Polushina
Faculty of Medicine and Dentistry, Department of Clinical Science, University of Bergen, Postboks 7804, NO-5020 Bergen, Norway
e-mail: Tatiana.Polushina@k2.uib.no

C. Sun et al. (eds.), *Signal and Image Analysis for Biomedical and Life Sciences*,
Advances in Experimental Medicine and Biology 823,
DOI 10.1007/978-3-319-10984-8_3

3.1 Introduction

Change-point problems (or disorder problems, break-point problems) are used to model heterogeneity in sequences of observations. This is essential in order to understand the underlying properties of a process as a part of the statistical diagnosis of data. Primarily it serves the purpose of checking and validating the homogeneity assumption of the data, which is one of the main assumptions in statistical modelling. Thus, accounting for these changes facilitates more improved and reliable estimates for unknown parameters. This is an imperative step in statistical modelling directly associated with the decision making process. Change-point detection problem has received increasing attention due to these reasons and has attracted wide range of applications in many scientific streams. These change-point models are employed in health informatics, financial and economic data analysis, signal processing, oceanographic studies, quality control, surveillance analysis, etc.

In health informatics, detection and characterization of genomic structural variations are essential in identifying disease causing genes that have functional importance in exemplifying genome-wide complex diseases, such as cancer, autism, immune disorders, etc. These structural variations in the human genome can be acquired somatically in the lifespan as well as be inherited through germline. Copy number variation (CNV) is one of the common and major types of structural variations in the human genome. CNV is defined as a DNA segment that is 1 kb or larger and present at variable copy number in comparison with a reference genome [9]. It is identified in multiple studies that CNV plays an important role in genetic susceptibility to common diseases [27, 40]. There are multiple platforms and procedures built to detect CNV in different perspectives [3, 12, 20, 55]. The array comparative genomic hybridization (aCGH) is a popular and a widely used methodology to detect CNVs in genome-wide studies. It is developed on the principles of the conventional comparative genomic hybridization (CGH) technique [14], which produces a map of DNA sequence copy number with respect to the chromosomal location. The CGH technique was firstly developed to detect copy number changes in solid tumors. In CGH experiments, the differentially labeled test and control genomes are hybridized to metaphase chromosomes. The fluorescent signal intensity of the test DNA relative to the reference DNA along the chromosome is linearly plotted to identify CNVs. The aCGH technique uses slides arrayed with small segments of DNA as the targets for analysis [20] in contrast to the use of metaphase chromosomes in CGH. The aCGH technique offers high resolution for CNV detection. Moreover, simultaneously detection of different alterations types is one of the advantages of the CGH technique [51]. Furthermore, it has been proven that aCGH is a powerful tool for detecting submicroscopic chromosomal abnormalities in individuals with idiopathic mental retardation and various birth defects.

There is a large amount of literature on CNV detection in aCGH data. A method based on fitting a mixture of three Gaussian distributions corresponding to gain, loss and normal regions is considered in [11]. Later, a test based on moving averages proposed in [26] to compute a threshold level to detect CNVs. In [24], a modified version of the circular binary segmentation [41] introduced. Their methodology is termed as circular binary segmentation (CBS) method. A test based on the maximum of a likelihood ratio is used in the CBS to detect CNVs. The method discussed in [24] is employed in the popular *DNAcopy* R package [36, 53]. Different methods based on hidden Markov models (HMMs) introduced in [40, 46]. Furthermore, a fast Bayesian change-point detection method based on the product partition models [1] introduced in [7] and it is deployed in the *bcp* R package [6]. A different approach for the problem discussed in [22] which uses the "lars" algorithm [5] and a generalized version [54] of the BIC [39] to estimate change-points in aCGH data. The methodology is freely available in the *cumSeg* R package [21]. Recently, a Pruned Exact Linear Time (PELT) method is introduced in [17]. The *changepoint* R package [16] employs the methods discussed in [17]. Readers are referred to [19] for a detailed review on the segmentation methods on aCGH data.

Detection of CNVs falls into the posteriori (retrospective or off-line) class of change-point problems. In the posteriori change-point problem the data set is fixed and it is not getting changed periodically as in the sequential (quickest or on-line) change-point problem. There exists an extensive literature on both of these main classes of change-point problems. Readers are referred to [2,15,23,31,32,34,47,48] for a detailed review on some of the techniques. The quickest change-point problem, a sequence of random variables is observed on-line, that is, the future observations are not known. Initially, we assume that the sequence considered is in so-called "controlled" state. But at some unknown moment a breakage occurs and the sequence runs "out of control". The objective of sequential change-point analysis is to detect this breakage (change-point) as soon as possible with a minimum number of false alarms. There are two well-known sequential procedures discussed in the literature: the Shiryaev-Roberts (SR) procedure [37,42–44] and the Cumulative Sum (CUSUM) procedure [25].

The process of change-point analysis in both the retrospective and sequential change-point methods deals with two main issues: detecting number of change-points and estimating their locations. In this chapter, we propose novel hybrid algorithms that combine sequential change-point techniques and the Cross-Entropy (CE), which is a model based stochastic optimization technique. We emphasize that the hybrid algorithm in [31] is based on a genetic algorithm and a local search procedure, whereas the proposed method uses sequential change-point techniques and the CE algorithm. Our method utilizes a sequential change-point detection methodology to provide initial estimates on the number as well as the locations of the change-points. Based on the initial estimates the CE method is initiated to optimize the solution to provide more accurate estimates of the number as well as their corresponding locations. We propose two new algorithms within

this framework. The first approach, which combines the SR procedure and the CE method, will be referred as the "SR-CE". The second approach combines the CUSUM procedure and the CE method. We will refer to this method as the "CUSUM-CE". In this study, we apply the proposed algorithms to aCGH data in order to detect CNVs. Notice that the new hybrid algorithms can easily be extended or modified to solve change-point problems in other research fields.

This chapter is organized as follows. First, we describe the multiple change-point problem. Then we provide details on the proposed hybrid algorithms, quickest change-point detection methods and the CE method. In the numerical results section, we present results on simulated data and two publicly available real datasets. Finally, in the discussion and conclusions section, we consider the strengths and limitations of the proposed methodology and conclude the chapter with future research directions.

3.2 Multiple Change-Point Problem

Let us consider a sequence of observations $\mathbf{X} = (x_1, x_2, \ldots, x_L)$ of length L, in which the x_i's are independently distributed Gaussian random variables. A segmentation of the sequence is specified by the number of change-points N and the corresponding locations of the change-points $\mathbf{C} = (c_1, c_2, \ldots, c_N)$, where $1 = c_0 < c_1 < \cdots < c_N < c_{N+1} = L + 1$. A change-point is defined as a boundary between two adjacent segments in this context. The value of c_i is the sequence position of the rightmost character of the segment to the left of the i-th change-point. The segments are numbered from 0 to N as there will be one or more segments than the number of change-points. The model assumes that within each segment the observations are distributed as normal with mean μ_i, $i = 0, 1, \ldots, N$ and variance σ^2. Both mean and variance are not known in advance and maximum likelihood method is used to obtain estimates. The joint distribution of $\mathbf{x}$ conditional on N, $\mathbf{C} = (c_1, c_2, \ldots, c_N)$, $\boldsymbol{\mu} = (\mu_0, \mu_1, \ldots, \mu_N)$, and σ^2 is given by:

$$f(\mathbf{X} \mid N, \mathbf{C}, \boldsymbol{\mu}, \sigma^2) = \prod_{n=0}^{N} \left[\prod_{i=c_n}^{c_{n+1}-1} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{ -\frac{(x_i - \mu_n)^2}{2\sigma^2} \right\} \right]. \quad (3.1)$$

The corresponding log-likelihood of the model is

$$ll(\mathbf{X} \mid N, \mathbf{C}, \boldsymbol{\mu}, \sigma^2) = \sum_{n=0}^{N} \left[-\frac{\lambda_n}{2} \ln(2\pi\sigma^2) - \frac{1}{2} \sum_{i=c_n}^{c_{n+1}-1} \left(\frac{x_i - \mu_n}{\sigma} \right)^2 \right], \quad (3.2)$$

where the length of the n-th segment is defined as $\lambda_n = c_{n+1} - c_n$.

3.3 Framework of the Algorithms

The proposed algorithms combine a sequential procedure with the CE method to detect multiple change-points in biological sequences of continuous measurements. We consider both the SR procedure and CUSUM procedure to combine with the CE method to form a hybrid framework to detect multiple change-points. In general, the SR-CE and CUSUM-CE hybrid algorithms can be summarized as follows:

1. Run a sequential procedure (either SR or CUSUM) along the sequence of observations to obtain initial estimates for the number (N) as well as the locations ($\mathbf{C}$) of change-points.
2. Based on the estimates of N and $\mathbf{C}$, initiate the CE algorithm to obtain an optimized locations of change-points.
3. For all pairs of adjacent segments, perform a two sample t-test to identify the least important change-point that associated with the highest p-value with respect to the significance level (α). The p-values are adjusted on the Bonferroni correction [45] to control the family wise error rate in multiple hypothesis testing. Thus, we eliminate the least significant change-point from the solution and update the solution vector with the other estimates.
4. Initiate the CE algorithm with the new set of change-point locations.
5. Repeat steps 3 and 4 until all change-points found are significant. Return $\mathbf{C}$: the vector of change-point locations. The length of this vector is the number of change-points.

3.3.1 Quickest Change-Point Detection

The sequential change-point problem can be described in mathematical terms as follows. Let $\{X_n\}_{n\geq 1}$ be independent random variables which are observed sequentially, one by one. Suppose that initially the sequence is in so-called "controlled" state for $n = 1, 2, \ldots, \tau - 1$, that is, the random variables are distributed with $f_0(x)$, a common normal probability density function with mean μ_0 and variance σ^2. At some unknown moment τ a breakage occurs and the observed sequence runs "out of control", which means that after the breakage (change-point) the probabilistic characteristics of the sequence are changed. From moment τ we observe random variables with $f_1(x)$, $f_1(x) \neq f_0(x)$, another normal probability density function with mean μ_1 and variance σ^2. Our objective is to detect the change-point as soon as feasible and with as few as possible false alarms. In other words, in the sequential change-point problem, we would like to detect the moment τ as quickly as possible after it has occurred and, at the same time, we would like to keep the rate of false alarms at a low predefined level.

There are two main cases in the sequential change-point problem [49]. In the simplest situation, we know the probability density functions before and after the breakage, which may be unrealistic assumption. In the second case, we assume that

the $f_0(x)$ is known before the change-point, whereas the $f_1(x)$ is unknown. In what follows we assume that the μ_0 is known (it can be estimated from an archive of data) and the μ_1 is unknown and must be estimated from the data, the σ^2 is fixed. For the sake of simplicity of the formulas below (and without loss of generality), we can assume that $\mu_0 = 0$ and $\sigma^2 = 1$.

We have two statistical hypotheses: the null hypothesis H_0: there is no change-point versus the alternative hypothesis H_1: a breakage happens at time $\tau = k \geq 0$. The sequential decision rule can be constructed as follows. Let $\mathbf{X}_n = (X_1, X_2, \ldots, X_n)$ be a vector of the first $n \geq 1$ values. The probability density functions of $\mathbf{X}_n$ under either of these hypotheses are given by

$$p(\mathbf{X}_n \mid H_0) = \prod_{j=1}^{n} f_0(X_j),$$

$$p(\mathbf{X}_n \mid H_1) = \prod_{j=1}^{k-1} f_0(X_j) \prod_{j=k}^{n} f_1(X_j), \quad k \leq n.$$

Then we can calculate the likelihood ratio, which can be used to test H_0 versus H_1, as follows

$$LR_k = \prod_{j=k}^{n} \frac{f_1(X_j)}{f_0(X_j)} = \prod_{j=k}^{n} \frac{\frac{1}{\sqrt{2\pi}} \exp\{-\frac{1}{2}(X_j - \hat{\mu}_1)^2\}}{\frac{1}{\sqrt{2\pi}} \exp\{-\frac{1}{2}X_j^2\}} = \exp\left\{ \frac{\left(\sum_{j=k}^{n} X_j\right)^2}{2(n-k+1)} \right\},$$

where $\hat{\mu}_1 = \sum_{j=k}^{n} X_j/(n-k+1)$, is an estimate of the μ_1 based on the last $n-k+1$ observations.

There are two common characteristics of a sequential detection procedure: the average run length (ARL) to false alarm (the expected number of values to an alarm assuming that there is no breakage) and the average delay to detection (the expected delay between a change and its detection). The objective is to find a sequential procedure that minimizes the average detection delay with restriction on the ARL to false alarm.

In this chapter, we consider two main procedures: the Shiryaev-Roberts (SR) procedure [37, 42–44] and the CUSUM procedure [25]. Various probabilistic properties of these methods are discussed in [28–30].

The SR procedure stops and raises an alarm at time

$$T_{A_{SR}} = \inf\{n \geq 1 : R_n \geq A_{SR}\},$$

where

$$R_n = \sum_{k=1}^{n} LR_k = \sum_{k=1}^{n} \exp\left\{ \frac{\left(\sum_{j=k}^{n} X_j\right)^2}{2(n-k+1)} \right\}, \quad n = 1, 2, \ldots$$

is the SR statistic, and A_{SR} is a positive threshold that controls the false alarm rate.

The stopping time of the CUSUM procedure is defined by

$$T_{A_C} = \inf\{n \geq 1 : W_n \geq A_C\},$$

where

$$W_n = \max_{1 \leq k \leq n} LR_k = \max_{1 \leq k \leq n} \exp\left\{ \frac{\left(\sum_{j=k}^{n} X_j\right)^2}{2(n-k+1)} \right\}, \quad n = 1, 2, \ldots$$

is the CUSUM statistic, and A_C is an unknown threshold that controls the false alarm rate in the CUSUM procedure.

In order to identify the thresholds A_{SR} and A_C we generate an artificial sequence with a single change-point and apply the SR and the CUSUM procedures for the sequence. We assume that the first several observations are in "controlled" state. Therefore, the estimates of the initial (unknown) parameters of the probability density function $f_0(x)$ can be obtained using these first observations. In this study, we consider random observations with normal distribution $\texttt{Normal}(\mu_0, 1)$ before the change-point and with $\texttt{Normal}(\mu_1, 1)$ after the change-point, where $\mu_0 = 0$, $\mu_1 = 1$. We use 2,000 as the length of the sequence and 100 as the number of observations utilized for estimating the parameters of the initial distribution. After simulating this experiment 2,000 times, it is clear that the threshold A_{SR} should be quite large (see Table 3.1). For instance, if $A_{SR} = 5{,}000$, then the probability of detecting the true change-point is 0.739.

Note that p, the probability of detecting the true change-point, increases as the threshold A_{SR} increases (see Fig. 3.1). We can conclude that for long sequences we should use large values of the threshold A_{SR}, for example, $A_{SR} > 5{,}000$, in order to detect the change-point with a high probability. Figure 3.2 shows the overall processing time(s) with respect to different threshold values, A_{SR}.

Since the estimates of the positions of the change-points will be used as the initial values for the CE method, we should emphasize that these estimates are found with some delay. Tables 3.2 and 3.3 show how the ARL depends on the value of $\mu_1 - \mu_0$. These tables also demonstrate that the ARL is significantly large for small differences between μ_1 and μ_0. Note for rather long sequences this delay is not an issue whereas for relatively short sequences we should use a higher value of the A_{SR} and reduce the length of the region used for estimating unknown parameters of $f_0(x)$.

Table 3.1 The average run length (ARL), CPU time and the probability of detecting the true change-point for different values of the threshold A_{SR} (Reprinted with permission from [35]. Copyright 2013, AIP Publishing LLC)

A	700	800	900	1,000	2,000	3,000	4,000	5,000	6,000	7,000
p	0.411	0.432	0.472	0.493	0.629	0.679	0.715	0.739	0.765	0.779
ARL	108	107	109.3	109	110.5	112.8	113.2	115.7	113.5	116.7
CPU time (s)	0.23	0.39	0.58	0.65	0.82	1.07	1.21	1.34	1.46	1.48

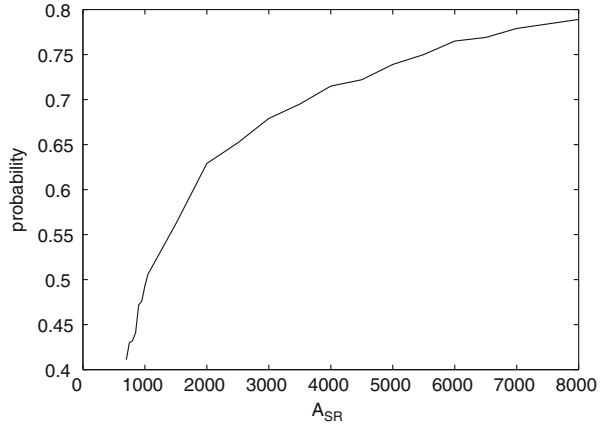


Fig. 3.1 The probability of detecting the true change-point depending on the value of the threshold A_{SR} (Reprinted with permission from [35]. Copyright 2013, AIP Publishing LLC)

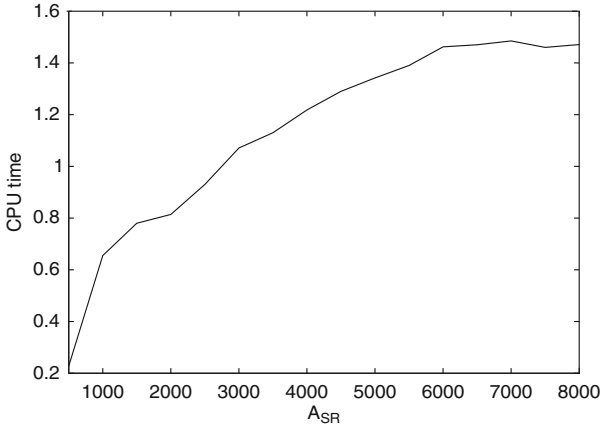


Fig. 3.2 CPU time depending on the value of the threshold A_{SR} (Reprinted with permission from [35]. Copyright 2013, AIP Publishing LLC)

Table 3.2 The ARL and CPU time for different values of $\mu_1 - \mu_0$, $A_{SR} = 5{,}000$ (Reprinted with permission from [35]. Copyright 2013, AIP Publishing LLC)

$\mu_1 - \mu_0$	0.25	0.5	0.75	1	1.25	1.5	1.75	2	5	7
ARL	324.9	162.9	125.7	115.7	109.5	103.7	103.0	102.4	99.2	95.0
CPU time (s)	2.10	1.94	1.65	1.39	1.13	1.02	0.92	0.94	0.87	0.85

Table 3.3 The ARL and CPU time for different values of $\mu_1 - \mu_0$, $A_{SR} = 7{,}000$ (Reprinted with permission from [35]. Copyright 2013, AIP Publishing LLC)

$\mu_1 - \mu_0$	0.25	0.5	0.75	1	1.25	1.5	1.75	2	5	7
ARL	351.8	172.7	129.7	116.7	108.7	107.4	106.2	105.5	102.1	99.0
CPU time (s)	2.47	2.03	1.82	1.46	1.31	1.27	1.15	0.98	0.94	0.88

Table 3.4 The average run length (ARL) and the probability of detecting the true change-point for different values of the threshold A_C (Reprinted with permission from [35]. Copyright 2013, AIP Publishing LLC)

A	1,000	5,000	7,000	10,000	12,000	15,000	18,000	19,000	20,000	21,000
p	0.269	0.583	0.642	0.672	0.701	0.694	0.728	0.743	0.752	0.748
ARL	101.1	101.7	111.4	112.6	110.5	112.3	114.3	114.2	115.5	115.4
CPU time (s)	0.63	1.17	1.26	6.047	1.34	1.54	1.48	1.48	1.49	1.53

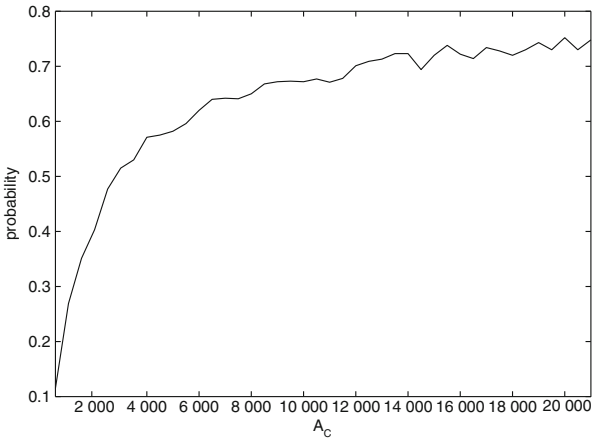


Fig. 3.3 The probability of detecting the true change-point depending on the value of the threshold A_C

We repeat the same simulation for the CUSUM procedure. Using the simulated sequences, we estimate the threshold A_C (note that A_C is significantly larger than A_{SR}) (see Table 3.4). For large values of A_C the probability of detecting the true change-point increases very slow. Figure 3.4 shows the overall processing time(s) with respect to different threshold values, A_C. It is customary that the value of the False Discovery Rate (FDR) less than 0.25 is used as a popular threshold [50]. We use higher values of A_C and A_{SR}, since the probabilities of detecting the true change-point in the both cases are larger than 0.75 (Figs. 3.1 and 3.3).

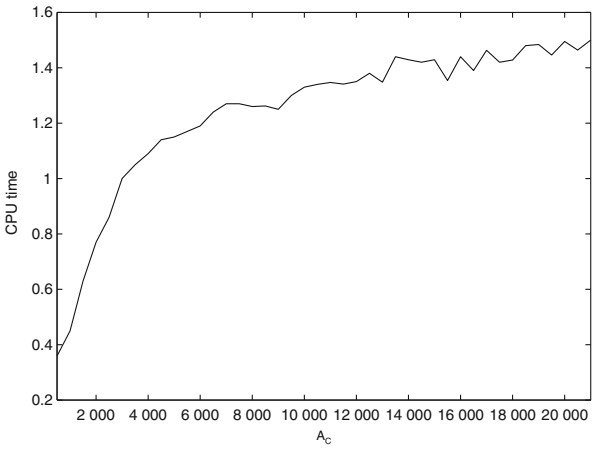


Fig. 3.4 CPU time depending on the value of the threshold A_C

3.3.2 The Cross-Entropy Method

The Cross-Entropy (CE) method [38] is a model-based evolutionary stochastic optimization framework which was originally developed as a method to estimate rare event probabilities. It can be used to solve both estimation and optimization problems. The CE method is developed on the basis of the Kullback-Leibler divergence [18]. The process of multiple change-point detection can be viewed as a combinatorial optimization problem. In the context of combinatorial optimization problems, the CE method is an iterative procedure that starts with a parametrized sampling distribution from which M number of random samples generated. Then, each combinatorial arrangement is scored for its performance using an objective function F. A fixed number of best performing combinatorial arrangements are selected based on the performance score and it is referred as the elite sample. We define the size of this elite sample as M_{elite}. Let us define $M_{elite} = \rho \times M$, where ρ is the elite sample fraction. The elite sample is used to update the parameters of the sampling distribution based on a smoothing rule. This process is iterated until a stopping criterion (SC) is met or user defined number of iterations. The sampling distribution eventually converges to a degenerate distribution about a locally optimal solution, which ideally will be globally optimal [4].

There are few parameters that have to be specified prior to the initialization of the CE method in the context of multiple change-point problem. They are the minimum aberration width (h), lower and the upper limit for the search space of number of change-points (let us define the lower limit as N_{min} and upper limit as N_{max}), sample size M, elite sample fraction ρ, smoothing parameter vector $\boldsymbol{\beta}$ and a cut-off value for the SC (ε). In this study, truncated normal distribution is utilized as the parametrized sampling distribution to simulate locations of the change-points based on the user defined minimum aberration width h. We simulate $M = 200$ number random solutions. The value of ρ is considered as 0.05, smoothing parameter values of β and

γ [8] is used for μ and σ respectively and ε is set as 0.01. The performance function used in the study is the model log-likelihood based on the simulated change-point locations.

Based on these user defined set of parameters and the initial estimates from a sequential method, the CE algorithm can be summarized as below:

1. Set the change-point locations obtained from a sequential procedure as the initial values for the mean vector $\boldsymbol{\mu^0}$ and set all components of the standard deviation vector $(\boldsymbol{\sigma}^2)^0$ as 5^2 in order to simulate locations from the truncated normal distribution. Both vectors of parameters are N-dimensional. Set $t = 0$.
2. Increase t by 1. Simulate a random sample $\mathbf{C}^{(1)}, \mathbf{C}^{(2)}, \ldots, \mathbf{C}^{(M)}$ from $\texttt{Normal}(\boldsymbol{\mu}^{t-1}, (\boldsymbol{\sigma}^2)^{t-1})$ distribution, where $\mathbf{C}^{(i)} = (c_1^{(i)}, c_2^{(i)}, \ldots, c_N^{(i)}), i = 1, 2, \ldots, M$.
3. For each $i = 1, 2, \ldots, M$ order $c_1^{(i)}, \ldots, c_N^{(i)}$ from smallest to largest and set $\mathbf{C}^{(i)} = (c_1^{(i)}, c_2^{(i)}, \ldots, c_N^{(i)})$, where $\mathbf{C}^{(i)}$ is the change-point vector as defined earlier.
4. Evaluate the log-likelihood function (the performance score) of each $\mathbf{C}^{(1)}, \mathbf{C}^{(2)}, \ldots, \mathbf{C}^{(M)}$. Obtain the elite sample, which is the best performing combinations of the change-point locations. M_{elite} is the size of the elite sample.
5. For all $j = 1, 2, \ldots, N$ calculate maximum likelihood estimates of the mean and the standard deviation $\hat{\boldsymbol{\mu}}^{\boldsymbol{t}} = (\hat{\mu}_1, \hat{\mu}_2, \ldots, \hat{\mu}_N)^t$, $(\tilde{\boldsymbol{\sigma}}^{\boldsymbol{2}})^t = (\tilde{\sigma}_1^2, \tilde{\sigma}_2^2, \ldots, \tilde{\sigma}_N^2)^t$ by using the elite sample. Based on the smoothing rule update the parameters in the truncated normal distribution as below,

$$\boldsymbol{\mu}^{\boldsymbol{t}} = \beta \hat{\boldsymbol{\mu}}^t + (1-\beta)\hat{\boldsymbol{\mu}}^{t-1}, \quad (\boldsymbol{\sigma}^2)^t = \gamma(\tilde{\boldsymbol{\sigma}}^2)^t + (1-\gamma)(\tilde{\boldsymbol{\sigma}}^2)^{t-1}.$$

6. If the stopping criterion (SC) is met, then stop the process and identify the combination of the locations of change-points $\mathbf{C}^{(i)}$ that optimizes the performance function. Otherwise set $t = t + 1$ and iterate from step 2. In this study, we use a SC based on the Mean Absolute Deviation (MAD) [10], which is a robust measurement on dispersion.

   SC is: Stop the process if $\max_j \text{MAD}_j < \varepsilon$, for all $j = 1, 2, \ldots, N$.

   where

$$\text{MAD}_j = \underset{i=1,2,\ldots,M}{\text{Median}} \left| c_j^{(i)} - \text{Median}\left(c_j^{(1)}, c_j^{(2)}, \ldots, c_j^{(M)}\right)\right|$$

   for all $j = 1, 2, \ldots, N$.

3.3.2.1 Bonferroni Correction for Multiple Hypothesis Testing

The Bonferroni correction is a conservative method that can be used to control the overall significance level (α) or the family wise error rate (FWER) when conducting multiple hypotheses tests. If $T_1, T_2, \ldots, T_n$ is a set of n statistics with corresponding p-values $P_1, P_2, \ldots, P_n$ for testing hypotheses $H_1, H_2, \ldots, H_n$, the

general Bonferroni multiple test procedure is performed by rejecting $H_i : i = 1, \ldots, n$ if the p-value (P_i) is less than or equals to α/n [45]. Thus, the Bonferroni inequality,

$$P\left\{\bigcup_{i=1}^{n}\left(P_i \leq \frac{\alpha}{n}\right)\right\} \leq \alpha \quad (0 \leq \alpha \leq 1)$$

ensures that the probability of rejecting at least one hypothesis when all are true is no greater than the significance level α, which is the type I error rate.

3.4 Numerical Results

We include results of numerical experiments to validate and assess the proposed hybrid algorithms. First, an artificially generated data set is considered with different signal-to-noise ratio (SNR) values as well as with different segment widths. The SNR is defined as the segment mean divided by the standard deviation of the Gaussian noise in the process as considered in [19]. Finally, two of the well-known publicly available real aCGH data sets are considered to further demonstrate the effectiveness of the proposed methodology.

In order to assess the performance of the proposed SR-CE and CUSUM-CE algorithms over the standard CE method [33] a comparison study is carried out; which is the primary focus of this study. The variant of the CE method discussed in this study utilizes a multi-core architecture based parallel implementation in the R statistical software [36]. Furthermore, for the completeness of the study, we compare the results obtained through the proposed methodology with another four well established change-point detection methods in the literature: *DNAcopy* [53], *bcp* [6], *changepoint* [16] and *cumSeg* [21]. In all of these methodologies we consider the default parameter values in the respective algorithms [19], as most user will be exercising.

3.4.1 Results on Artificially Generated Data

Let us consider a random sequence of length 3,500 with 10 abrupt change-points which results in having 11 segments. The standard deviation of the Gaussian noise is set as 1 in all the segments. Table 3.5 shows the parameter values used for the simulation study.

We follow the general work flow discussed in the framework of algorithms. First, based on a sequential procedure initial estimates of the change-point locations are obtained. Second, the CE algorithm is initiated based on these pre-estimates. We utilize the parameter values for the CE algorithm as described earlier with the smoothing parameters $\beta = 1$ and $\gamma = 1$. A significance level (α) of 0.001 [13]

Table 3.5 Parameter values for the simulation study

	Segment										
	1	2	3	4	5	6	7	8	9	10	11
Length	200	550	150	250.0	500	250	400	600	200.0	150.0	250
SNR	0	2	4	2.5	0	2	3	4	2.5	3.5	1
Mean[a]	0	2	4	2.5	0	2	3	4	2.5	3.5	1

[a]SNR=Mean/S.D., Standard Deviation is set as 1

Table 3.6 Initial estimates of locations and processing time of SR-CE and CUSUM-CE methods

Sequential procedure	Initial estimates for locations	Avg. proc. time (s)
SR	207, 422, 585, 755, 908, 1,058, 1,154, 1,509, 1,653, 1,910, 2,094, 2,310, 2,749, 2,914, 3,107, 3,205, 3,259	0.394
CUSUM	209, 755, 910, 1,154, 1,578, 1,654, 1,926, 2,308, 2,914, 3,107, 3,207, 3,262	0.923

is considered in the two sample t-test to assess the statistical significance of the identified change-points. The Bonferroni correction is used to control the family wise error rate in multiple hypothesis testing [45]. In the standard CE method, we set N_{min} as 1 and N_{max} as 20 as the search space for the number of change-points.

Table 3.6 shows the initial estimates for the locations obtained by the two sequential procedures. It is observed that both methods have over estimated the true number of change-points as expected, even though the processing time for both methods are less than a second. The mean profile plots of the original data, hybrid algorithms and the other methods are shown in Fig. 3.5. It is observed that the proposed SR-CE and CUSUM-CE procedures have correctly identified the true number of change-points ($N = 10$) as compared to the marginal over estimation ($N = 12$) given by the CE method. Except for the *cumSeg* method all other methods have over-estimated number of change-points, where the former method has under estimated the number of change-points. We observed that the over-estimation problem in *changepoint* and in *bcp* is severe than the other competing methods. However, when considering the average root mean square error (RMSE) both the proposed procedures and the general CE method have effectively segmented data with almost overlapping mean profiles. Table 3.7 shows the summary statistics on the performance of the proposed methods with the standard CE method.

Both of the proposed methods and the CE method have resulted with lower average RMSE rates. The SR-CE gives the lowest average RMSE rate with an approximate average improvement of 10 % over the CE method. The notable performance achievement in the proposed procedures when compared to the CE method is on the overall processing time. It is observed that both SR-CE and CUSUM-CE have performed significantly better than the general CE method, which even utilizes a multi-core architecture based parallel implementation. Among the two proposed methods, CUSUM-CE procedure performs better than the SR-CE

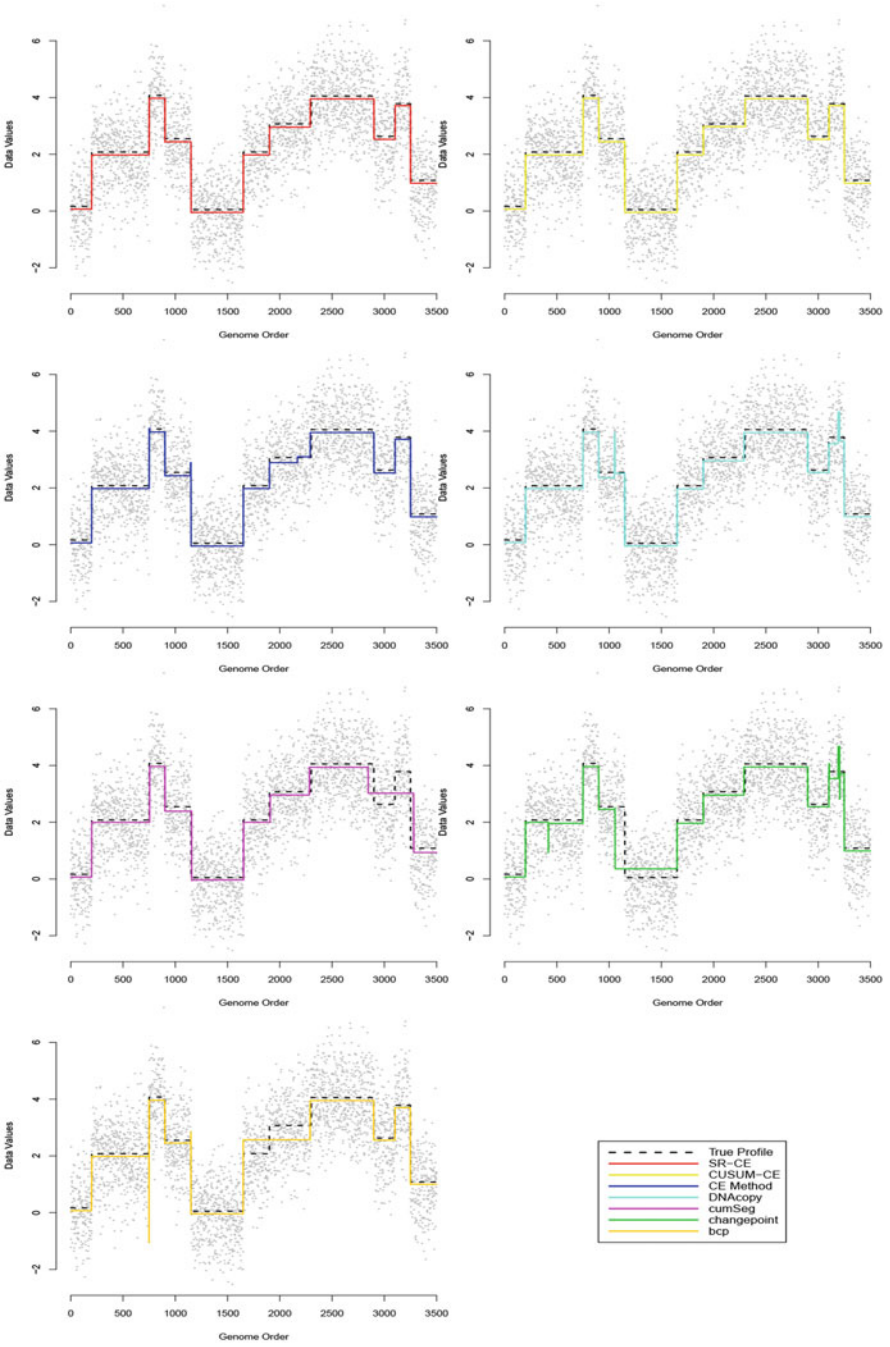


Fig. 3.5 Mean profile plots of the proposed algorithms and the other methods for the artificial data

Table 3.7 Summary statistics on the performance of hybrid frameworks and the CE method

	SR-CE	CUSUM-CE	CE method
Average RMSE	0.083	0.094	0.092
Median processing time (s)[a]	19.868	11.419	37.977

[a]Relative to a 2.3 GHz Intel Core i7 processor (Mac OS X 10.9) with 8 GB RAM.

Table 3.8 Initial estimates of change-point locations for the chromosomes 1, 3 and 7 of the GM03563 cell line data

	Chromosome		
Procedure	1	3	7
SR	122	57	131
CUSUM	81	59	161

when considering the overall processing time. The SR-CE procedure gives a significant improvement of around 91 %, while CUSUM-CE the improvement is more than two-folds (233 %) as compared to the processing time of the CE method.

3.4.2 Results on Real Data

3.4.2.1 Fibroblast Cell Lines Data

This example considers a cDNA microarray-based CGH data of fibroblast cell lines which was originally discussed in [46]. The data set it is freely available to download at http://www.nature.com/ng/journal/v29/n3/suppinfo/ng754 _S1.html. The data set consists of a single experiment on 15 fibroblast cell lines and it has already been discussed by several authors [22, 53] in the literature. We analyze the data in the fibroblast cell line GM03563 with respect to the chromosomes 1, 3 and 7. By spectral karyotyping, real alteration (a single change-point) is only found in chromosome 3 out of the considered chromosomes.

By utilizing the SR and CUSUM procedures, we obtain initial estimates of the change-point locations for the chromosomes 1, 3, 7 as in Table 3.8 separately. We initiate the CE algorithm with the same set of parameters considered in the artificial data example for all three chromosomal level data of GM03563 cell line. The standard CE method is initiated with the default parameter set with $N_{min} = 1$ and $N_{max} = 10$. Figure 3.6 shows the array CGH profiles for the three chromosomes based on the proposed methods as well as the other competing methods. In chromosome 1, both the proposed hybrid methods have correctly identified the true number of change-points as "zero", as opposed to a single change-point estimation given by the CE method. For chromosomes 3 and 7, both the proposed algorithms and the CE method have given the same results estimating the correct number as well as the locations of the change-points. Considering the performances of the other methods, all other methods except for the *changepoint* have failed to estimate the number of change-points as zero for the chromosome 1. In fact all of them have over-estimated the true number. In general, the *bcp* method tends to over-estimate the true-number of change-points.

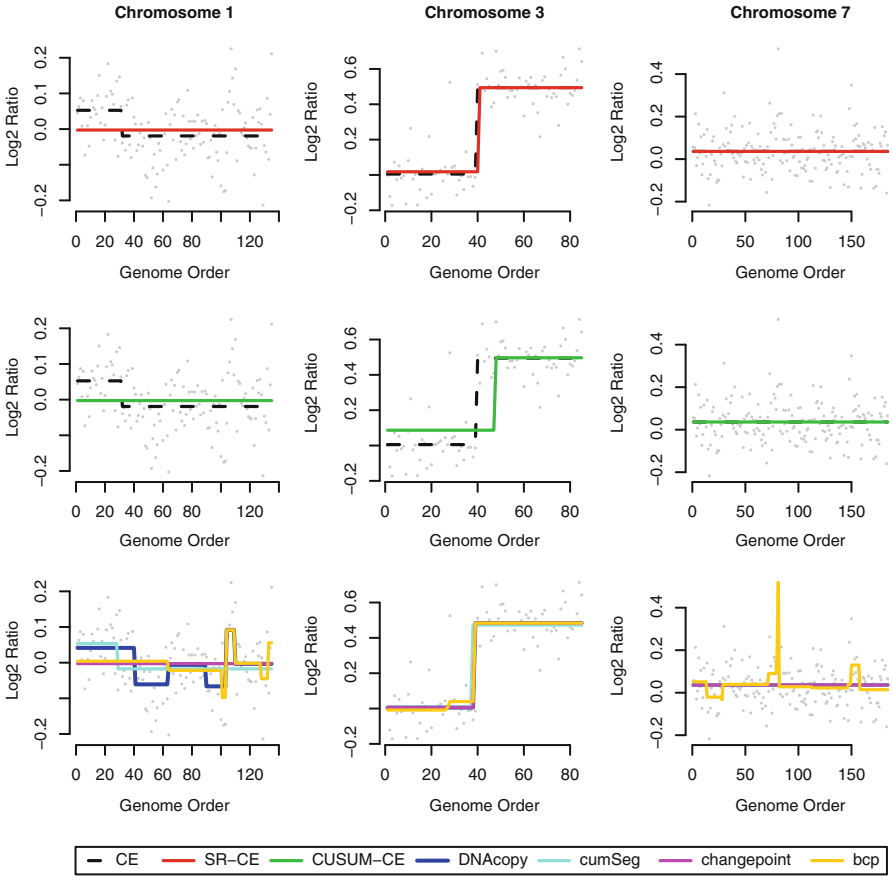


Fig. 3.6 Array CGH profiles for the chromosomes 1, 3 and 7 in GM03563 cell line

Table 3.9 Processing time (s) for the GM03563 cell line data

	Chromosome		
Method	1	3	7
SR-CE	0.079	0.126	0.112
CUSUM-CE	0.124	0.072	0.069
CE	2.533	2.435	3.151

Table 3.9 shows the overall processing time. It is observed that both the proposed procedures are highly computationally efficient than the CE method. Furthermore, on average CUSUM-CE procedure is faster than the SR-CE procedure.

Table 3.10 Initial estimates of change-point locations for the chromosomes 3, 5, 9 and 13 of MDA157 cell line data

	Chromosome			
Procedure	3	5	9	13
SR	83, 174, 280, 387	84, 180, 250, 328	126, 181	97
CUSUM	82, 146, 326	60, 144, 234, 328	129, 169	86

Table 3.11 Processing time (s) for MDA157 cell line data

	Chromosome			
Method	3	5	9	13
SR-CE	1.145	1.108	0.184	0.071
CUSUM-CE	0.643	0.994	0.251	0.071
CE	4.426	3.922	4.462	3.036

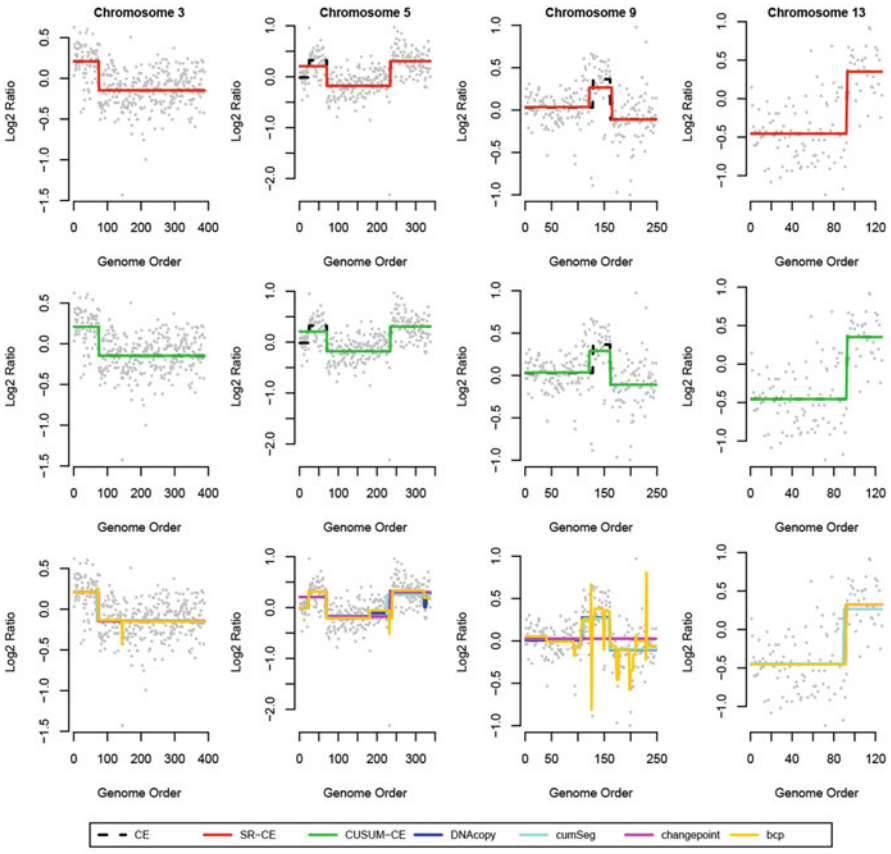


Fig. 3.7 Array CGH profiles for the chromosomes 3, 5, 9 and 13 in MDA157 cell line

3.4.2.2 Breast Tumor Data

In this example we consider the breast cancer cell line (MDA157) data which was originally discussed in [27]. The cDNA microarray CGH was profiled across 6,691 mapped human genes in 44 breast tumor samples and 10 breast cancer cell lines. This dataset is discussed in the [52] and [22] and can be downloaded from http://www.pnas.org/content/99/20/12963/suppl/DC1.

We apply our proposed algorithms as well as the other methods on chromosomes 3, 5, 9 and 13 data to estimate the underlying copy number variations. Initial estimates for the locations are given in Table 3.10 and overall processing time (in seconds) is given in Table 3.11. Figure 3.7 shows the aCGH profile plots for all the chromosomes. We observe that the proposed two procedures have behaved in a similar way in all chromosomes. Also, except for the chromosome 5, in all other cases CE method has also performed similar to SR-CE and CUSUM-CE procedures. In chromosome 9, *changepoint* method has not detected any change-points, whereas *bcp* has highly over-estimated the number of change-points. In general, our methods have similar profiles to *cumSeg* and *DNAcopy* procedures.

Discussions and Conclusions

We have proposed two novel hybrid algorithms (SR-CE, CUSUM-CE) that utilize powerful sequential change-point detection techniques (SR and CUSUM procedures) and a model based stochastic optimization technique (CE method) to estimate both the number and the locations of change-points in biological data of continuous measurements. This is the first-of-its kind implementation in the change-point literature that utilize on-line change point detection techniques to obtain initial estimates for a posteriori change-point problem and merge them with a model based stochastic optimization method (CE) to further improve the estimates on both the number and their corresponding locations. We compare the performance of the proposed hybrid algorithms with the standard CE algorithm, which does not use results from the sequential techniques as an input. Furthermore, for the completeness we have further compared our procedures with four other established change-point techniques. The effectiveness of the proposed methodology is assessed both in terms of artificially generated and real data. In all of the studies, it was found that the hybrid methods perform significantly better than the standard CE method both in terms of the precision and the processing time. In the standard CE method processing time is considered as one of the drawbacks in its implementation. Thus, incorporating sequential techniques has solved not only this critical issue for a greater extent, but also it has improved the detection power as well. Furthermore, use of the sequential techniques provide an upper limit for the search space for the number of change-points in the CE method. In the standard CE method user has to define these lower and upper search limit unknowingly. Thus, sequential techniques provide an

(continued)

important support for the standard CE method to perform more efficiently. While the results of this work are encouraging, there are plenty of avenues available as future research directions. In our study, it was identified that the sequential procedure is sensitive to the aberration width (i.e., segment width) resulting to favour analysis of longer sequences over the shorter sequences. Therefore, a versatile implementation of the methodology is worth for probing, which will work effectively in short as well as long sequences of data. Finally, the proposed procedures only developed to detect changes in mean levels of continuous measurements. We hope to extend these procedures to detect changes in the variance as well.

Acknowledgements W. J. R. M. Priyadarshana acknowledges the funding received towards his PhD from the International Macquarie University Research Excellence (iMQRES) scholarship. The authors acknowledge the anonymous referees for their useful comments.

References

1. D. Barry, J.A. Hartigan, A Bayesian analysis for change point problems. J. Am. Stat. Assoc. **88**, 309–319 (1993)
2. J.V. Braun, H.G. Müller, Statistical methods for DNA sequence segmentation. Stat. Sci. **13**, 142–162 (1998)
3. N.P. Carter, Methods and strategies for analyzing copy number variation using DNA microarrays. Nat. Genet. **39**, S16–S21 (2007)
4. A, Costa, O.D. Jones, D. Kroese, Convergence properties of the cross-entropy method for discrete optimization. Oper. Res. Lett. **35**, 573–580 (2007)
5. B. Efron, T. Hastie, I. Johnstone, R. Tibshirani. Least angle regression. Ann. Stat. **32**, 407–451 (2004)
6. C. Erdman, J.W. Emerson, bcp: an R package for performing a Bayesian analysis of change point problems. J. Stat. Softw. **23**, 1–13 (2007)
7. C. Erdman, J.W. Emerson, A fast Bayesian change point analysis for the segmentation of microarray data. Bioinformatics **24**(19), 2143–2148 (2008)
8. G.E. Evans, G.Y. Sofronov, J.M. Keith, D.P. Kroese, Identifying change-points in biological sequences via the cross-entropy method. Ann. Oper. Res. **189**, 155–165 (2011)
9. L. Feuk, A.R. Carson, S.W. Scherer, Structural variation in the human genome. Nat. Rev. Genet. **7**(2), 85–97 (2006)
10. D.C. Hoaglin, F. Mosteller, J.W. Tukey, *Understanding Robust and Exploratory Data Analysis* (Wiley, New York, 1983)
11. G. Hodgson, J.H. Hager, S. Volik, S. Hariono, M. Wernick, D. Moore, N. Nowak, D.G. Albertson, D. Pinkel, C. Collins, D. Hanahan, J.W. Gray, Genome scanning with array CGH delineates regional alterations in mouse islet carcinomas. Nat. Genet. **29**, 459–464 (2001)
12. S. Ivakhno, T. Royce, A.J. Cox, D.J. Evers, R.K. Cheetham, S. Tavare, CNAseg-a novel framework for identification of copy number changes in cancer from second-generation sequencing data. Bioinformatics. **26**, 3051–3058 (2010)
13. V.E. Johnson, Revised standards for statistical evidence. Proc. Natl. Acad. Sci. (2013). doi:10.1073/pnas.1313476110

14. A. Kallioniemi, O.P. Kallioniemi, D. Sudar, D. Rutovitz, J.W. Gray, F. Waldman, D. Pinkel, Comparative genomic hybridization for molecular cytogenetic analysis of solid tumors. Science **258**, 818–821 (1992)
15. J.M. Keith, Segmenting eukaryotic genomes with the generalized Gibbs sampler. J. Comput. Biol. **13**, 1369–1383 (2006)
16. R. Killick, I. Eckley, changepoint: an R package for changepoint analysis. R package version 1.1. (2013). http://CRAN.R-project.org/package=changepoint
17. R. Killick, P. Fearnhead, I. Eckley, Optimal detection of changepoints with a linear computational cost. J. Am. Stat. Assoc. **107**, 590–598 (2012)
18. S. Kullback, R.A. Leibler, On information and sufficiency. Ann. Math. Stat. **22**, 79–86 (1951)
19. W.R. Lai, M.D. Johnson, R. Kucherlapati, P.J. Park, Comparative analysis of algorithms for identifying amplifications and deletions in array CGH data. Bioinformatics **21**, 3763–3770 (2005)
20. R. Lucito, J. Healy, J. Alexander, A. Reiner, D. Esposito, M. Chi, L. Rodgers, A. Brady, J. Sebat, J. Troge, J.A. West, S. Rostan, K.C.Q. Nguyen, S. Powers, K.Q. Ye, A. Olshen, E. Venkatraman, L. Norton, M. Wigler, Representational oligonucleotide microarray analysis: a high-resolution method to detect genome copy number variation. Genome Res. **13**, 2291–2305 (2003)
21. V.M.R. Muggeo, cumSeg: change point detection in genomic sequences. R package version 1.1. (2012). http://CRAN.R-project.org/package=cumSeg
22. M.R.V. Muggeo, G. Adelfio, Efficient change point detection for genomic sequences of continuous measurements. Bioinformatics **27**, 161–166 (2011)
23. J. Oliver, P. Bernaola-Galvan, P. Carpena, R. Roman-Roldan, Isochore chromosome maps of eukaryotic genomes. Gene **276**(1–2), 47–56 (2001)
24. A.B. Olshen, E.S. Venkatraman, R. Lucito, M. Wigler, Circular binary segmentation for the analysis of array-based DNA copy number data. Biostatistics, **5**, 557–572 (2004)
25. E.S. Page, Continuous inspection schemes. Biometrika **41**, 100–115 (1954)
26. J.R. Pollack, C.M. Perou, A.A. Alizadeh, M.B. Eisen, A. Pergamenschikov, C.F. Williams, S.S. Jeffrey, D. Botstein, P.O. Brown, Genome-wide analysis of DNA copy-number changes using cDNA microarrays. Nat. Genet. **23**(1), 41–46 (1999)
27. J.R. Pollack, T. Sørlie, C.M. Perou, C.A. Rees, S.S. Jeffrey, P.E. Lonning, R. Tibshirani, D. Bo, D. Botstein, A.L. Børresen-Dale, P.O. Brown, Microarray analysis reveals a major direct role of DNA copy number alteration in the transcriptional program of human breast tumors. Proc. Natl. Acad. Sci. U.S.A. **99**, 12963–12968 (2002)
28. M. Pollak, A.G. Tartakovsky, Exact optimality of the Shiryaev-Roberts procedure for detecting changes in distributions, in *Information Theory and its Applications, ISITA 2008 International Symposium*, Auckland (2008), pp. 1–6
29. M. Pollak, A.G. Tartakovsky, Optimality properties of the Shiryaev-Roberts procedure. Statistica Sinica **19**, 1729–1739 (2009)
30. A. Polunchenko, G. Sokolov, W. Du, Quickest change-point detection: a bird's eye view, in *Joint Statistical Meeting (JSM)*, Montreal (2013)
31. T. Polushina, G. Sofronov, Change-point detection in biological sequences via genetic algorithm, in *Proceedings IEEE Congress on Evolutionary Computation (CEC)*, New Orleans (2011), pp. 1966–1971
32. T.V. Polushina, G.Y. Sofronov, A hybrid genetic algorithm for change-point detection in binary biomolecular sequences, in *Proceedings of the IASTED International Conference on Artificial Intelligence and Applications (AIA 2013)*, Innsbruck (2013), pp. 1–8
33. W.J.R.M. Priyadarshana, G. Sofronov, A modified cross entropy method for detecting multiple change points in DNA Count Data, in *WCCI 2012 IEEE World Congress on Computational Intelligence (CEC)*, Brisbane (2012), pp. 1020–1027
34. W.J.R.M. Priyadarshana, G. Sofronov, GAMLSS and extended cross-entropy method to detect multiple change-points in DNA read count data, in *Proceedings of the 28th International Workshop on Statistical Modelling*, Palermo, vol. 1, ed. by V.M.R. Muggeo, V. Capursi, G. Boscaino, G. Lovison (2013), pp. 453–457

35. W.J.R.M. Priyadarshana, T. Polushina, G. Sofronov, A hybrid algorithm for multiple change-point detection in continuous measurements, in *International Symposium on Computational Models for Life Sciences*, Sydney, ed. by C. Sun, T. Bednarz, T.D. Pham, P. Vallotton, D. Wang. AIP Conference Proceedings (2013), pp. 108–117
36. R Core Team:R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing (2013), http://www.R-project.org/
37. S.W. Roberts, A comparison of some control chart procedures. Technometrics **8**, 411–430 (1966)
38. R. Rubinstein, D.P. Kroese, *The Cross-Entropy Method: a Unified Approach to Combinatorial Optimization, Monte-Carlo Simulation and Machine Learning* (Springer, New York, 2004)
39. G. Schwarz, Estimating the dimension of a model. Ann. Stat. **6**, 461–464 (1978)
40. J. Sebat, B. Lakshmi, J. Troge, J. Alexander, J. Young, P. Lundin, S. Maner, H. Massa, M. Walker, M. Chi, N. Navin, R. Lucito, J. Healy, J. Hicks, K. Ye, A. Reiner, T.C. Gilliam, B. Trask, N. Patterson, A. Zetterberg, M. Wigler. Large-scale copy number polymorphism in the human genome. Science **305**, 525–528 (2004)
41. A. Sen, M. Srivastava, On tests for detecting a change in mean. Ann. Stat. **3**, 98–108 (1975)
42. A.N. Shiryaev, The problem of the most rapid detection of a disturbance in a stationary process. Soviet Mathmatics. Dokl. **2**, 795–799 (1961)
43. A.N. Shiryaev, On optimum methods in quickest detection problems. Theory Probab. Appl. **8**, 22–46 (1963)
44. A.N. Shiryaev, *Optimal Stopping Rules* (Springer, New York, 1978)
45. R.J. Simes, An improved Bonferroni procedure for multiple tests of significance. Biometrika **73**, 751–754 (1986)
46. A.M. Snijders, N. Nowak, R. Segraves, S. Blackwood, N. Brown, J. Conroy, G. Hamilton, A.K. Hindle, B. Huey, K. Kimura, S. Law, K. Myamboo, J. Palmer, B. Ylstra, J.P. Yue, J.W. Gray, A.N. Jain, D. Pinkel, D.G. Albertson, Assembly of microarrays for genome-wide measurement of DNA copy number. Nat. Genet. **29**, 263–264 (2001)
47. G. Sofronov, Change-point modelling in biological sequences via the Bayesian adaptive independent sampler. Int. Proc. Comput. Sci. Inf. Technol. **5**, 122–126 (2011)
48. G.Y. Sofronov, G.E. Evans, J.M. Keith, D.P. Kroese, Identifying change-points in biological sequences via sequential importance sampling. Environ. Model. Assess. **14**, 577–584 (2009)
49. G. Sofronov, T. Polushina, W.J.R.M. Priyadarshana, Sequential change-point detection via the cross-entropy method, in *The 11th Symposium on Neural Network Applications in Electrical Engineering (NEUREL'12)*, Belgrade (2012), pp. 185–188
50. A. Subramanian, H. Kuehn, J. Gould, P. Tamayo, J.P. Mesirov, GSEA-P: a desktop application for Gene set enrichment analysis. Bioinformatics **23**, 3251–3253 (2007)
51. A. Theisen, Microarray-based comparative genomic hybridization (aCGH). Nat. Educ. **1**(1), 45 (2008)
52. R. Tibshirani, P. Wang, Spatial smoothing and hot spot detection for CGH data using the fused lasso. Biostatistics **9**, 18–29 (2008)
53. E.S. Venkatraman, A. Olshen, DNAcopy: DNA copy number data analysis. R package version 1.34.0 (2013)
54. H. Wang, B. Li, C. Leng, Shrinkage tuning parameter selection with a diverging number of parameters. J. R. Stat. Soc. Ser. B (Stat. Methodol.) **71**, 671–683 (2009)
55. C. Xie, M.T. Tammi, CNV-seq, a new method to detect copy number variation using high-throughput sequencing. BMC Bioinformatics **10**, 80 (2009)

Chapter 4
Stochastic Anomaly Detection in Eye-Tracking Data for Quantification of Motor Symptoms in Parkinson's Disease

Daniel Jansson, Alexander Medvedev, Hans Axelson, and Dag Nyholm

Abstract Two methods for distinguishing between healthy controls and patients diagnosed with Parkinson's disease by means of recorded smooth pursuit eye movements are presented and evaluated. Both methods are based on the principles of stochastic anomaly detection and make use of orthogonal series approximation for probability distribution estimation. The first method relies on the identification of a Wiener model of the smooth pursuit system and attempts to find statistically significant differences between the estimated parameters in healthy controls and patients with Parkinson's disease. The second method applies the same statistical method to distinguish between the gaze trajectories of healthy and Parkinson subjects tracking visual stimuli. Both methods show promising results, where healthy controls and patients with Parkinson's disease are effectively separated in terms of the considered metric. The results are preliminary because of the small number of participating test subjects, but they are indicative of the potential of the presented methods as diagnosing or staging tools for Parkinson's disease.

Keywords Smooth pursuit • Parkinson's disease • Parametric modeling • Nonparametric modeling • Visual stimulus design • Eye tracking

D. Jansson (✉) • A. Medvedev
Department of Information Technology, Uppsala University, Uppsala, Sweden
e-mail: daniel.jansson@it.uu.se; am@it.uu.se

H. Axelson
Department of Neuroscience, Neurophysiology, Uppsala University, Uppsala, Sweden
e-mail: hans.axelson@akademiska.se

D. Nyholm
Department of Neuroscience, Neurology, Uppsala University, Uppsala, Sweden
e-mail: dag.nyholm@neuro.uu.se

C. Sun et al. (eds.), *Signal and Image Analysis for Biomedical and Life Sciences*,
Advances in Experimental Medicine and Biology 823,
DOI 10.1007/978-3-319-10984-8_4

4.1 Introduction

This chapter deals with mathematical modeling and identification of the human smooth pursuit system (SPS) and the application of the models to motor symptom quantification in Parkinson's disease (PD).

The SPS is a complex neuromuscular system governing smooth pursuit eye movements (SPEM), and the task is to keep a moving target in the visual field [16, 21]. Attempting to initiate smooth pursuit without a moving visual stimulus is difficult and usually results in a series of saccades that are swift simultaneous movements of both eyes. Saccades and smooth pursuit are the two ways in which humans can shift gaze [3].

The oculomotor system, and thus also the SPS, is impaired in a wide variety of neurological diseases [1, 5, 14, 16, 19]. In particular, PD is known to be associated with deficit in SPEM control [16]. Diagnosing and quantifying the disease is done by interview and clinical observation which requires hours of interaction between the patient and a qualified clinician [18]. Acquiring a better understanding of the SPS cast in mathematical models may be a first step towards developing a technology that allows for fast and automatic PD staging.

Studying the SPS requires means for eye movement measurement. The two most common techniques for eye movement registration are electrooculography (EOG) and video eye tracking. In EOG, electrodes are placed around the eye to measure the potential differences produced by the retina as it turns. Assuming that the resting potential is constant, the recorded potential is a measure of the eye's position. Video eye tracking uses one or more cameras, usually infrared, together with image analysis algorithms to locate the pupils and to determine the gaze direction. Video eye tracking is non-invasive, but the resolution is lower than in EOG.

Lately, the increased performance and accessibility of eye tracking technologies have generated a great deal of interest in the commercial sector. Examples are found in market research, gaze-based interaction, sports education etc. However, most applications use the eye tracker output directly and in a static manner, with interest only in where the subject is looking at a given time. This chapter presents an effort towards developing more sophisticated data analysis techniques in an attempt to extract previously hidden information from the eye tracking data and to open up for new more advanced applications.

The SPS relates gaze direction to visual stimuli and may thus be viewed as a dynamical system with an input and an output signal [7]. Modeling any dynamical system can be done in one of three ways: white-box modeling, grey-box modeling, or black-box modeling.

White-box modeling requires complete knowledge of the system to be modeled. The models are based on first principles and derived from physical laws. Although very useful, white-box models are relatively uncommon due to the exceeding complexity of most processes in nature.

Grey-box modeling requires partial knowledge of the system to be modeled. Through certain insight into the system, a semi-physical description of it can be

obtained, where one or more of the model parameters have been assigned physical meaning. Grey-box models are important tools and are widely used to predict and evaluate the behavior of countless processes in a range of industrial and scientific applications.

In black-box modeling, a general model structure is assumed without any physical meaning in the parameters. There is a number of common black-box structures for both linear and nonlinear systems. Black-box models provide simple means of evaluating the dynamics of a system at low computational cost and without the need for any deeper understanding of its nature.

In the case of the SPS, white-box modeling is a near impossible task due to the vast complexity of the feedback loop; the interaction between the eyes and the brain, which is affected by the not easily modeled human consciousness. Instead, this chapter considers parametric and non-parametric black-box models to portray the SPS.

This chapter will first present brief overviews of the extraocular system, SPEM, eye-tracking techniques, and Parkinson's disease to provide a background for the following modeling and experimental work. Then, some probability theory basics, including probability density estimation, are given, followed by descriptions of the two methods for SPS characterization. A way of designing sufficiently exciting visual input signals is then derived and the experiment setup is presented. Finally, the chapter is concluded by some results and a discussion.

4.2 The Extraocular Muscles

There are six muscles governing the movement of the eye, referred to as the *extraocular muscles*. Four of the muscles control the movement of the eye in the four directions: up, down, left and right. The remaining two muscles control the adjustments in gaze direction involved in counteracting head movements. The four muscles controlling standard eye movement are the superior, inferior, lateral and medial recti. The two remaining muscles are the superior and inferior oblique. The primary action of the superior and inferior recti are elevation (upward movement) and depression (downward movement) respectively and those of the lateral and medial recti are abduction (away from the median sagittal plane of the body) and adduction (towards the sagittal plane). Figure 4.1 shows the right eye with its accompanying extraocular muscles.

4.3 Smooth Pursuit

The two ways in which humans can voluntarily shift gaze are SPEM and saccades [16]. Saccades are discrete movements that quickly change the orientation of the eyes, thereby translating the image of the object of interest from an eccentric

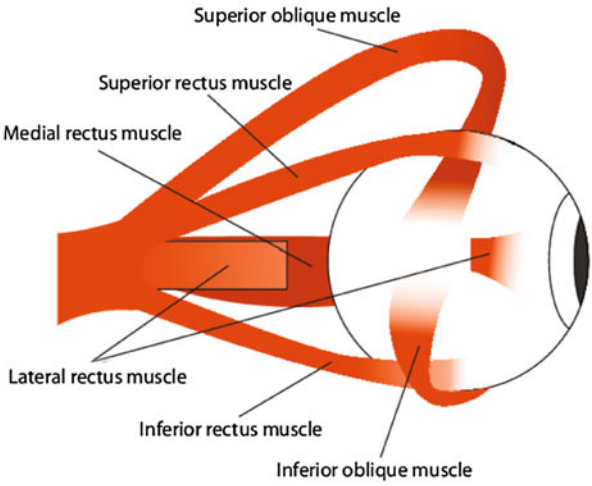


Fig. 4.1 The extraocular muscles of the right eye

retinal location to the fovea (the center of the retina responsible for sharp central vision). Smooth pursuit is a continuous movement that slowly rotates the eyes to track the motion of an object and to keep it within the visual field. SPEM are governed by SPS. Smooth pursuit is primarily driven by visual motion which makes it difficult for most individuals to initiate it without a moving target. The maximum angular velocity of the eyes during smooth pursuit is about 80–100 °/s [17]. For targets with exceeding velocities, the SPS passes the control to the saccadic system. Research has shown that direction-selective, motion-sensitive cells in the primary visual cortex estimate target angular velocity [19] and that the SPS acts as a velocity servo; in that it tries to minimize the angular velocity error between the gaze and the target [21]. Any stationary error in angular position will be left uncorrected by this mechanism.

There are several research papers on quantifying the SPS in an attempt to use SPEM as a biometric [2, 13]. Most papers on the subject are published in medical journals and apply straightforward and facile techniques for data analysis. The Smooth Pursuit Gain (SPG), the ratio between the eye velocity and the stimuli velocity, has been used as a measure for characterizing the SPS [14, 16]. Since the SPG is nothing but the steady-state angular velocity gain, it is merely one point in the frequency response of the SPS and may thus not be an exhaustive metric.

4.4 Eye Tracking

Eye tracking is the process of measuring either the point of gaze or the motion of the eye relative to the head. The two most common techniques for eye movement registration are EOG and video eye tracking. In EOG, electrodes are placed around the eye to measure the potential differences produced by the retina as it

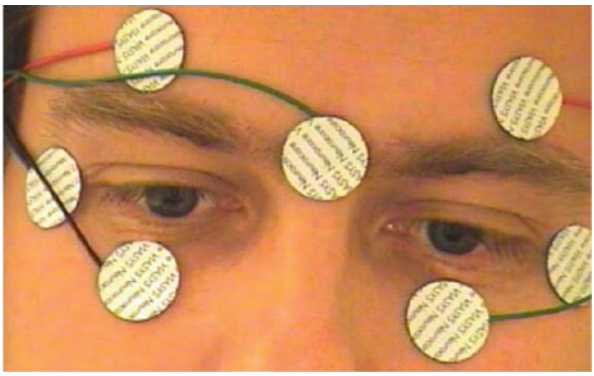


Fig. 4.2 EOG electrodes placed around the eyes to measure the potential differences produced when the eye turns

turns (see Fig. 4.2). Assuming that the resting potential is constant, the recorded potential is a measure of the eye's angular position. The signal acquired from an EOG measurement is called the electrooculogram. Because the EOG relies on the potential differences produced by a shift in the angular position of the retina, it is possible to use EOG even when the eyes are closed and it can thus be used in sleep studies. A drawback of the EOG is the fact that the resting potential is often not constant, resulting in nonlinear trends in the recorded data. Another drawback is the somewhat daunting task of placing the electrodes which also induces some discomfort in the test subject due to the need for a thorough scrubbing where the electrodes are to be placed.

Video eye tracking uses one or more cameras, usually infrared, together with image analysis algorithms to locate the pupils via the corneal reflections and to determine the gaze direction. Video eye tracking is non-invasive and quick, but a simple calibration procedure of the individual is needed before using the eye tracker. When more than one camera is utilized, the images from the different cameras can be combined to form a 3D environment, allowing for accurate tracking of the position and orientation of the head, which greatly improves the gaze direction measurements.

In the presented experiment, a two-camera video eye tracking system from Smart Eye AB, Sweden is used. The eye tracker output yields the number of centimeters (horizontal and vertical components separately) from the monitor center to the point where the gaze vector intersects the monitor. The system samples the gaze direction at 60 Hz. Figure 4.3 is a screenshot from the eye tracking software showing how the algorithms have found the gaze direction and orientation of the head.

4.5 Parkinson's Disease

PD is a degenerative disorder of the central nervous system. The cause of the disease is attributed to degeneration of dopaminergic neurons from the substantia nigra [12]. Impairment of the substantia nigra to synthesize dopamine causes a progressive

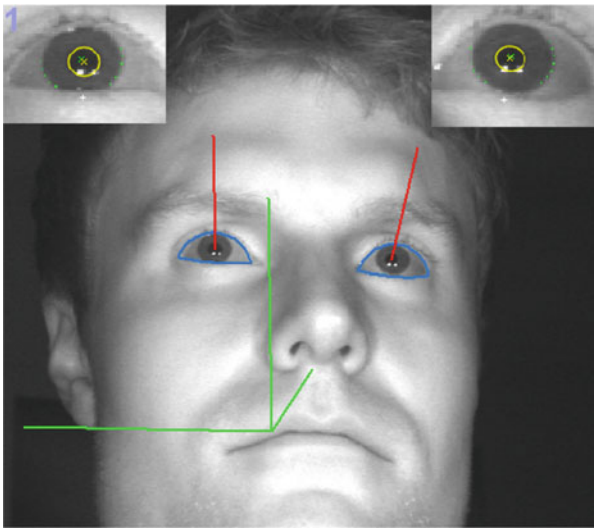


Fig. 4.3 A screenshot from the eye tracking software

depletion of this significant neurotransmitter for the putamen and caudate nucleus. The progression of PD is characterized by tremor during rest, abnormal gait, muscular rigidity and impaired balance [12].

Currently, the status of PD in a patient is evaluated through the Unified Parkinson's Disease Rating Scale (UPDRS) that is qualitatively interpreted by a clinician [20]. Knowledge of the current status of the disease in a patient is important for the selection and dosage of drug therapy. There are two major issues with the scale. The process of observing and interviewing the patient to determine the UPDRS result is time-consuming and often tiring for both the patient and the clinician. Moreover, since the scale is qualitatively interpreted there may be variation among the subjective decisions of different clinicians. Hence, it is of great interest to find means for quick and objective quantification of the PD status in a patient.

It has been shown that the SPS is negatively affected by PD and that the severity of the impairment is related to the progression of the disease [16, 18, 26]. Consequently, acquiring a full understanding of the SPS may be a first step towards developing a technology that allows for fast and automatic PD staging.

4.6 Probability Density Estimation

Probability density function (PDF) estimation from data is a broad topic and several different techniques exist. The most straightforward PDF estimator is the histogram that requires a relatively large sample size to yield a good approximation.

Two other methods are used in this chapter: kernel density estimation (KDE) which approximates the density function by a normalized sum of kernel functions, and an orthogonal series approximation (OSA) approach where the density function is expanded in terms of an orthonormal basis.

4.6.1 Stochastic Variables

Given a stochastic variable X of dimension M with distribution D, a sample from D is a set of observations $\{x_i\}_{i=1}^N$, where $x_i \in \mathbb{R}^M$ are realizations of X. From the sample, information about the underlying distribution can be extracted. For instance, the sample mean

$$\hat{\mu} = \frac{1}{N}\sum_{i=1}^{N} x_i, \tag{4.1}$$

and covariance

$$\hat{\Sigma} = \frac{1}{N-1}\sum_{i=1}^{N}(x_i - \hat{\mu})(x_i - \hat{\mu})^T, \tag{4.2}$$

are unbiased estimators of the true mean and covariance of the distribution.

The PDF of the distribution, denoted by $f(x)$, may also be estimated from the sample, however this is a more demanding task.

4.6.2 Kernel Density Estimation

The kernel approximation of the PDF $f(x)$ is given by

$$\hat{f}(x) = \frac{1}{N|H|^{1/2}}\sum_{i=1}^{N}\phi(H^{-1/2}(x - x_i)), \tag{4.3}$$

where ϕ is the chosen kernel function which is symmetric and integrates to one [24]. The parameter $H \in \mathbb{R}^{M\times M}$ is known as the bandwidth of the kernel, and it is symmetric, positive definite, and acts as a smoothing parameter. If H is equal to hI, where I is the identity matrix, a high value of h will give a smooth estimate, with low variance but high bias. Conversely, a low value of h gives higher variance but a low bias of the estimate. Choosing H appropriately for a specific application must usually be done through experiment.

4.6.3 Orthogonal Series Approximation

The KDE method has the drawback that the approximation requires N terms, which is often a large number. An alternative to the kernel estimator is the orthogonal series estimator [22, 23] which has the capability of capturing the shape of the PDF using far fewer terms than the kernel estimator. Assume that X is supported in the domain Q, i.e., $P(X \in Q) = 1$. If f is square integrable ($f \in \mathbf{L}^2(Q)$), the density may be approximated with any desired accuracy by a truncated orthogonal series

$$\hat{f}(x) = \sum_{j \in J} c_j \varphi_j(x), \quad x \in S, \tag{4.4}$$

where

$$c_j = \int_Q f(x)\varphi_j(x)\mathrm{d}x, \tag{4.5}$$

J is a finite set of M-tuples of integers and $\{\varphi_j(x), j = 0, \ldots, \infty\}$ is an orthonormal basis. The largest integer in each dimension in J gives the highest order of that dimension and must be chosen by the user. The highest order in each dimension decides the number of the basis functions that will be used in the approximation. Note that because f is a probability density, the coefficients in the above mentioned partial sum can each be written as the expectation

$$c_j = \int_Q f(x)\varphi_j(x)\mathrm{d}x = E\{\varphi_j(X)\}. \tag{4.6}$$

Hence, estimating c_j can be done via the sample mean

$$\hat{c}_j = \frac{1}{N_s} \sum_{i=0}^{N_s} \varphi_j(x_i), \tag{4.7}$$

where $(x_1, x_2, \ldots, x_{N_s})$, $x_i \in \mathbb{R}^M$ are observations of the underlying stochastic variable.

Here, all considered distributions are of dimension two and the orthonormal basis is chosen to be

$$\{\varphi_{n_1, n_2}(x)\} = \left\{ \sqrt{\det(\Gamma)}\phi_{n_1,n_2}\left(\Gamma(x - \mu)\right) \right\}, \tag{4.8}$$

where $\{\phi_{n_1, n_2}\}$ is the complete set of two-dimensional Hermite functions. The vector $\mu \in \mathbb{R}^N$ and the matrix $\Gamma \in \mathbb{R}^{N \times N}$ are user parameters for scaling and translating the functions.

The Hermit functions, $\{\phi_{n_1, n_2}(x)\}$, $x \in \mathbb{R}^2$, are given by

$$\phi_{n_1,\,n_2}(x) = \frac{1}{\sqrt{2^{n_1}2^{n_2}n_1!n_2!\pi}} e^{-(x^T x)/2} \tilde{H}_{n_1,\,n_2}(x), \tag{4.9}$$

where

$$\tilde{H}_{n_1,\,n_2}(x) = 2^{\frac{n_1+n_2}{2}} H_{n_1,\,n_2}(\sqrt{2}x) \tag{4.10}$$

are the *physicists'* Hermite polynomials [15] and

$$H_{n_1,\,n_2}(x) = (-1)^{n_1+n_2} e^{(x^T x)/2} \frac{\partial^{n_1}}{\partial x_1^{n_1}} \frac{\partial^{n_2}}{\partial x_2^{n_2}} e^{-(x^T x)/2} \tag{4.11}$$

are the *probabilists'* Hermite polynomials [15].

The choice of the user parameters μ and Γ will depend on the data. The vector μ is chosen to be the sample mean of the observations x_i, i.e.,

$$\mu = \frac{1}{N_s} \sum_{i=1}^{N_s} x_i.$$

This choice of μ reduces the number of the Hermit functions required in the truncated series to achieve a given estimation error [22]. If Γ is chosen as a diagonal matrix, its diagonal elements will decide the width of the functions in the corresponding dimension. The choice of the diagonal elements should be based on the variance of the considered observations in each dimension. Choosing the functions to be too narrow will increase the required function order for accurate estimation, and choosing the functions to be too wide will smudge the estimated distribution, reducing the significance of single observations.

4.6.4 Finding the Outlier Region

Assume that an observation, $x \in \mathbb{R}^M$, is made and that it must be determined whether it is likely to be an observation of a given random variable X, or not. A hypothesis test with the null hypothesis:

- H_0: x is an observation of X

must be carried out. One way to do this is to define an *outlier region* S of the random variable, being the set of all possible observations deemed unlikely to come from the considered distribution, i.e., all x for which H_0 is rejected. The probability that an observation of X lies in S should be low. Define α such that

$$P(X \in S) = \int_S f(x)\mathrm{d}x = \alpha. \tag{4.12}$$

Consequently, α is the probability with which an observation of the considered random variable is deemed (incorrectly) to be from some other distribution. The choice of α will influence the size of the outlier region S.

A method for checking whether a given observation is an outlier (lies in S) is derived and described in detail in [11], but will only be outlined here. Finding the outlier region, S of a PDF f, can be done numerically by carrying out the following steps:

- Evaluate f for the finite set of uniformly spaced grid points $\{x_i\}_{i=1}^{L}$ to obtain $\{f_i\}_{i=1}^{L}$.
- Let $\{f_{(i)}\}_{i=1}^{L}$ be $\{f_i\}_{i=1}^{L}$ sorted in ascending order.
- Find K such that $\sum_{i=1}^{K} f_{(i)} \leq \frac{\alpha}{A} < \sum_{i=0}^{K+1} f_{(i)}$, where A is the area of a grid element.
- An approximation $\hat{S}$ of S is then given by

$$\hat{S} = \{x_i : f(x_i) \leq f_{(K)} = \gamma_T\}. \tag{4.13}$$

4.7 Non-parametric Method

The non-parametric method for distinguishing between individuals on the basis of their eye movements comprises two parts. First, by presenting the same stimulus trajectory to a test subject multiple times, a set of output trajectories is obtained. The trajectories can be seen as observations of a trajectory distribution, the PDF of which can be estimated from data. Then, independently acquired datasets, possibly from a different individual, are tested against the estimated distribution. The test determines whether or not the distribution and the recorded data pertain to the same individual.

Assume that N_s datasets of eye movements are recorded from a test subject tracking the same trajectory of the visual stimulus multiple times on different occasions. Due to the complex nature of the oculomotor system, the response to a visual stimulus will not be the same for repeated exposures. Hence, the N_s datasets will not be equal. For each of the N_t time instances at which the gaze direction is sampled, there will be N_s data points, one from each set of recorded eye movements. Since horizontal and vertical gaze direction coordinates are logged separately, the data points will have two components.

The data points at time instance k will be seemingly random with some expected value, and can thus be seen as N_s observations of a two-dimensional stochastic variable, $X(k)$. Note that there will be one stochastic variable per time instance. The distribution of $X(k)$ for each k can be estimated from data.

The distribution of $X(k)$ will depend on the trajectory of the visual stimulus, but also on the individual tracking it. If the PDF of $X(k)$ for $k = 1, 2, \ldots, N_t$ is known for an individual, i.e., if an eye tracking profile has been established for that person, it is possible to determine whether a given dataset is likely to come from the same

subject or not. For each of the N_t time instances, a hypothesis test with the null hypothesis that the data are indeed observations of $X(k)$ can be carried out. If the number of time instances at which the deviation of the data from the considered distribution is high, the dataset is deemed to not be from the considered eye tracking profile. The approach generalizes in a straightforward manner to the case of a group of test subjects sharing a property, such as healthy persons or persons of a certain age.

In practice, the distributions of $X(k)$, $k = 1, 2, \ldots, N_t$, are not known, but can be estimated from data. The simplest way is to use the histogram. However, since the data are two-dimensional, a large number of data points is needed to achieve sufficiently small bin widths for reliable statistical testing. To acquire a large number of data points, a test subject would have to track the same visual stimulus a large number of times, which would be time-consuming and tedious. Therefore, OSA and KDE are used for PDF estimation instead.

In this chapter, the described method is used to distinguish healthy test subjects from test subjects with Parkinson's disease. PDFs constituting the trajectory distribution for healthy subjects are estimated and tested against eye movement data of subjects with Parkinson's disease.

4.8 Parametric Method

In the parametric method, the SPS is modeled as a dynamical system. By identifying the model for several datasets from healthy individuals, the distribution of the model parameters, θ, can be estimated using KDE or OSA. Once a PDF for the values of θ characteristic of healthy subjects has been established, independently acquired parameter values can be tested against it. Values of θ deviating from the PDF indicate that the corresponding test subject is unhealthy.

Mathematical modeling of the SPS is a complicated task. The grey-box model used in [9] is accurate and its parameters have physical and biological meaning. However, due to the model complexity, identification is computationally demanding and theoretical evaluation, such as the investigation of statistical stationarity and stability, is difficult.

One important feature of the present work is SPS input design. It is desirable to obtain a model that is easy to work with, but still captures the overall behavior of the system adequately. A linear black-box model may seem as an appropriate first candidate for the task. However, as is revealed in [9], linear models are unable to accurately predict the amplitudes of the SPS output. A nonlinearity must be included in the model in order for the model output to comply with data. A simple approach to alleviate this problem is to augment the linear model with a static output nonlinearity to yield a Wiener model. The Wiener model consists of a linear dynamical block in cascade with a static nonlinear function as shown in Fig. 4.4. The SPS is a multiple-input-multiple-output (MIMO) system because both the input and output consist of horizontal and vertical components. For simplicity, the SPS is here assumed to

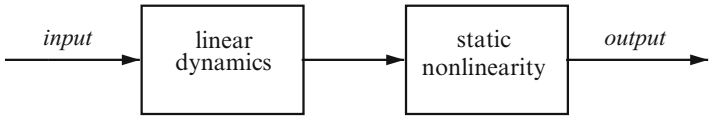


Fig. 4.4 The structure of a Wiener system

be represented by two single-input-single-output (SISO) systems in parallel, one governing the horizontal and one the vertical movements of the eyes, and thus neglecting any coupling between the two.

Here, the linear block is assumed to be a time-invariant ARX-structure given by

$$(1 + a_1 q^{-1} + a_2 q^{-2} + a_3 q^{-3}) x(n) = b q^{-4} s(n) + e(n), \tag{4.14}$$

where $x(n)$ is the output at time nT_s, $s(n)$ is the input at time nT_s, $e(n)$ is zero-mean white Gaussian noise with variance σ^2, $\theta = [a_1, a_2, a_3, b]^T$ is the parameter vector of the model, q is the forward time shift operator, and T_s is the sampling time. The output of the linear block is fed to a static nonlinearity, which in this study is for simplicity chosen to be a continuous piecewise-linear function. The resulting Wiener model of the horizontal or vertical components of the SPS is given by

$$\begin{cases} (1 + a_1 q^{-1} + a_2 q^{-2} + a_3 q^{-3}) x(n) = b q^{-4} s(n) + e(n), \\ y(n) = f\left(x(n)\right), \end{cases} \tag{4.15}$$

where $y(n)$ is the gaze direction at time nT_s and $f(x)$ is the static nonlinearity. For details on the choice of the linear dynamics and the static nonlinearity, please refer to [8]. The unknown parameters to be estimated in this model are θ and the parameters of f.

Identification of the model is carried out using the methods presented in [27].

4.9 Visual Stimuli

Visual stimuli are generated to have sufficient spectral and amplitude excitation for accurate identification of model (4.15). In practice, the stimuli are presented to the test subject as smooth random movements of a white circle in a $25 \times 25\,\mathrm{cm}^2$ black background window on a computer monitor. Horizontal and vertical components of each stimulus are generated separately as real-valued sequences. To ensure sufficient temporal excitation for identification of the dynamical part of (4.15), the sequences are required to have a desired (Fourier) spectral density d. Identification of the static nonlinearity in (4.15) requires the input to contain information for all possible amplitudes within the dynamic range, in this case limited by the size of the computer window. Sufficient amplitude excitation is more difficult to impose. The problem is addressed in [8] by constraining the sequences to have a

certain peak-to-average-ratio (PAR). If the vector 2-norm of s, denoted by $\|\cdot\|$, is constrained to be equal to $\sqrt{N}$, the PAR can be written as

$$\mathbf{PAR}(s) = \frac{\max_n |s_n|^2}{\frac{1}{N}\sum_{n=1}^{N} |s_n|^2} = \max_n |s_n|^2. \tag{4.16}$$

With $\|s\|^2 = N$, the upper bound for the PAR is N. The sequence s representing each component (horizontal and vertical) of the visual stimulus is generated as follows.

Let $s(t)$ be the sought continuous input signal over the time interval $[0, T_p]$. Discretisize the interval in N steps by $0 = t_0 < t_1 < \cdots < t_{N-1} < T_p$. The discrete version of the signal is then $s_n = s(t_n), n = 0, \ldots, N-1$. Let $s = [s_0\ s_1\ \ldots\ s_{N-1}]^T$ be the signal written in a vector form. Assume that there is some desired power spectral density (PSD) $d = [d_0\ d_1\ \ldots\ d_{N_f-1}]^T$ that the spectrum of s has to be as close to as possible in the least-squares sense. It is also desired for s to have a maximum allowed PAR.

The following optimization problem can be formulated to obtain an s with the two properties mentioned above:

$$\begin{aligned} \min_s &= \sum_{p=0}^{N-1} ||z_p| - \sqrt{d_p}|^2 \\ &\text{s.t. } \|s\|^2 = N \\ &\quad\ \mathbf{PAR}(s) \le \rho \end{aligned} \tag{4.17}$$

where ρ is some maximum allowable PAR and $z = [z_0\ z_1\ \ldots\ z_{N-1}]^T$ is the Fourier domain representation of s which can also be written as $z = \mathbf{F}^H s$, where $\cdot^H$ denotes the conjugate transpose operator. $\mathbf{F}^H$ is the Fourier matrix given by

$$\mathbf{F}^H = \frac{1}{\sqrt{N}} \begin{bmatrix} 1 & 1 & \ldots & 1 \\ 1 & \omega & \ldots & \omega^{N-1} \\ 1 & \omega^2 & \ldots & \omega^{2(N-1)} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & \omega^{N-1} & \ldots & \omega^{(N-1)(N-1)} \end{bmatrix}_{N \times N} \tag{4.18}$$

with $\omega = e^{-2\pi i/N}$. The minimization problem expressed by (4.17) is equivalent to

$$\begin{aligned} \min_{s,\varphi} &\sum_{p=0}^{N-1} |z_p - \sqrt{d_p} e^{i\varphi_p}|^2 \\ &\text{s.t. } \|s\|^2 = N \\ &\quad\ \mathbf{PAR}(s) \le \rho \end{aligned} \tag{4.19}$$

where $\varphi = [\varphi_0\ \varphi_1\ \ldots\ \varphi_{N-1}]^T$ is a vector of auxiliary variables and $\varphi_n \in [-\pi, \pi]$, as proven in [6]. Rewriting the problem in a more convenient vector form yields

$$\min_{s,\varphi} f = \|\mathbf{F}^H s - v\|^2 \tag{4.20}$$

$$\begin{aligned} \text{s.t. } & \|s\|^2 = N \\ & \mathbf{PAR}(s) \leq \rho, \end{aligned} \tag{4.21}$$

where $v = [\sqrt{d_0}e^{i\varphi_0}\ \sqrt{d_1}e^{i\varphi_1}\ \ldots\ \sqrt{d_{N-1}}e^{i\varphi_{N-1}}]$.

The minimization problem in (4.20) can be solved in a cyclic way. Fix s to any real sequence and compute the v that minimizes f. The value of v that minimizes f for a fixed s is obtained by letting

$$\varphi_p = \arg(\text{the } p^{th} \text{ element of } \mathbf{F}^H s), \tag{4.22}$$

where $\arg(\cdot)$ denotes the complex argument. Next fix v and write the minimization problem as

$$\min_{s} \|s - \mathbf{F}v\|^2 \tag{4.23}$$

$$\begin{aligned} \text{s.t. } & \|s\|^2 = N \\ & \mathbf{PAR}(s) \leq \rho \end{aligned} \tag{4.24}$$

where the fact that $\mathbf{F}^H\mathbf{F} = \mathbf{I}$, where $\mathbf{I}$ is the unitary matrix, was used to rewrite the criterion function. Problem (4.23) has been called the "nearest-vector" problem and can be solved using the methodology in [25]. To disregard the PAR constraint, ρ can be set to N, which is the maximum possible PAR for any sequence with the norm equal to $\sqrt{N}$. Iterating between (4.22) and (4.23) until convergence gives a signal s satisfying the design specifications. The design procedure is outlined in Algorithm 1. The algorithm has the property of monotonically decreasing the criterion as the iteration proceeds [4].

Algorithm 1 *Minimization of the criterion in* (4.20)

1. Initialize s to a random vector s^0. Iterate Step 2 and Step 3 below, for $i = 0, 1, \ldots$ until convergence.
2. $\varphi^i = \arg\min \mathbf{F}^H s^i$
3. Let s^{i+1} be the solution to the nearest-vector problem solved in [25] and set $i \leftarrow i + 1$.

end

Selecting d and ρ is not straightforward and appropriate values must be determined experimentally. Here, the spectral density d is chosen so that s has a fixed amount of energy for all frequencies below a certain cut-off frequency f_c, and no energy for frequencies above it. A larger f_c implies a stimulus with higher

average velocity. Mathematically, this choice of d can be expressed with the help of the sampling frequency, f_s, and the number of samples, N as

$$d_p = \begin{cases} \frac{f_s}{2f_c} & \text{if } 0 \le p < N\frac{f_c}{f_s} \\ 0 & \text{if } N\frac{f_c}{f_s} \le p < \frac{N(f_s-f_c)}{f_s} \\ \frac{f_s}{2f_c} & \text{if } \frac{N(f_s-f_c)}{f_s} \le p \le N \end{cases}. \tag{4.25}$$

The sequence d is made symmetric around its center in order for the corresponding time domain sequence to be real-valued. The constant value of the spectrum for frequencies smaller than f_c is chosen so that the total signal energy is indeed N, as was required in (4.17).

The maximum allowable PAR, ρ, is chosen to give an amplitude distribution of s that is as close to a uniform one as possible over the dynamic range.

Problem (4.17) is solved two times with different initial values of s to yield independent horizontal and vertical components of the stimulus. The two components are then combined to form the two-dimensional visual input.

4.10 Experiment

Gaze direction data of test subjects attempting to track the moving circle on a computer monitor were recorded using a video-based eye tracker from Smart Eye AB, Sweden. Test subjects were placed 50 cm from the monitor with the monitor center at eye height. The eye tracker output is the distance in centimeters (horizontal and vertical components separately) between the monitor center and the point where the gaze direction line intersects the plane of the monitor. Eye-tracking data were sampled at a sampling frequency of $f_s = 60$ Hz.

Stimuli of length $N = 1{,}560$ samples (about 26 s) were generated using the presented stimulus generation method with $f_c = 1$ Hz and $\rho = N/10$. The stimuli were then displayed to test subjects of different ages. The conducted experiment involved four healthy test subjects:

- H1: Man, 26 years old
- H2: Man, 27 years old
- H3: Man, 54 years old
- H4: Man, 64 years old

and five test subjects diagnosed with Parkinson's disease:

- P1: Woman, 57 years old
- P2: Man, 71 years old
- P3: Woman, 73 years old
- P4: Man, 66 years old
- P5: Man, 67 years old

4.11 Results

4.11.1 Non-parametric Method

Two-dimensional PDFs describing gaze direction at each time instant were estimated from 20 datasets of both H1 and H3 using OSA. Datasets from all test subjects were then used to compare with the estimated distributions, i.e., to test whether they were observations from the distributions in question or not. The datasets of H1 and H3 used to test against the estimated distributions were independent and not utilized in the OSA estimation.

For the estimation of OSA, the user parameter μ in (4.8) was chosen to be the sample mean of the observations. Γ in (4.8) was selected to be a diagonal matrix with diagonal elements $r_i = \frac{2}{\hat{\sigma}_i}, i = 1, 2$, where $\hat{\sigma}_i$ is the sample standard deviation in each dimension of the considered sample of observations, i.e., of the horizontal and vertical data in each time instant. This choice of Γ was made experimentally and shown to give the most satisfactory results. The highest order of the Hermit functions was set to four in each dimension. Using higher orders showed little or no improvement in the obtained results.

A heat map over an excerpt of the estimated trajectory distribution of H1 overlaid by the gaze trajectory of P1 is depicted in Fig. 4.5. It is apparent that the trajectory of P1 deviates from the mean trajectory of H1 at several time instants.

Table 4.1 shows the mean number of outliers at the 0.05 significance level in the datasets of different test subjects when comparing to the distribution estimates of H1 and H3, respectively.

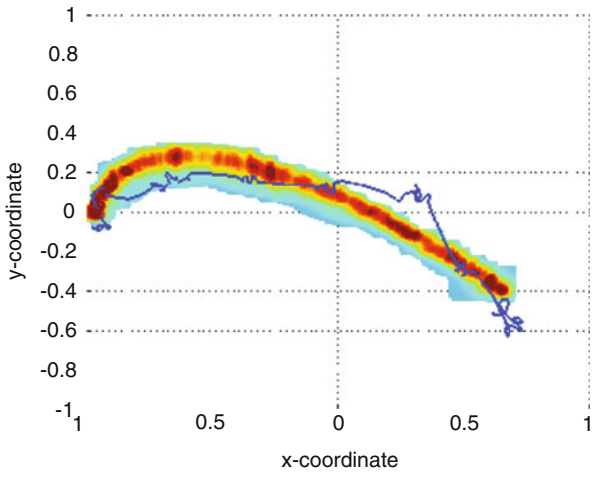


Fig. 4.5 Heat map of the estimated trajectory distribution of H1. *Red* indicates high values. The *blue line* shows a trajectory of P1 attempting to track the same stimulus (Reprinted with permission from [10]. Copyright 2013, AIP Publishing LLC)

Table 4.1 The average number of outliers, m, at the 0.05 significance level in the datasets of different test subjects when comparing to the distribution estimates of H1 and H3. The numbers are given as percent of the total number of samples in the dataset (Reprinted with permission from [10]. Copyright 2013, AIP Publishing LLC)

H1				H3			
Subject	m (%)	Subject	m (%)	Subject	m (%)	Subject	m (%)
P1	48.1	H1	0.6	P1	32.1	H1	11.7
P2	61.4	H2	14.3	P2	57.7	H2	8.1
P3	58.2	H3	26.3	P3	57.9	H3	1.1
P4	54.4	H4	39.6	P4	46.1	H4	20.9
P5	53.9			P5	46.6		

The entries in Table 4.1 give the amount of time during which the gaze direction of the test subjects deviated significantly from the mean trajectory of H1 and H3, respectively.

It is evident from Table 4.1 that the trajectories of the Parkinson patients deviate from the mean trajectories of the healthy subjects. It should also be noted that the trajectories of H1 and H3 show little deviation from the mean trajectories estimated from their own datasets, despite the fact that the sets used to test were not included in the distribution estimation.

4.11.2 Parametric Method

The parameters of (4.15) were estimated for the horizontal part of 10 eye-tracking datasets of each test subject. The parameter estimates were used together with OSA to approximate the distribution of the parameters in each test subject. In order to facilitate visual presentation, only the parameters a_1 and b were considered in the distribution estimation.

Figure 4.6 shows the boundary of the outlier regions of the distribution estimates. It can be seen that the parameters of Parkinson patients differ significantly from the corresponding parameters for the healthy controls.

Discussions and Conclusions

This chapter evaluates possibilities of distinguishing between patients diagnosed with Parkinson's disease and healthy controls on the basis of their recorded eye movements. The two suggested methods, one parametric and one non-parametric, rely on probability density function approximation and stochastic anomaly detection. The non-parametric method only shows the parts of a stimulus trajectory that deviate from a normal gaze profile established from data. The parametric method attempts to quantify the whole

(continued)

trajectory in terms of a few model parameters and can provide insights into the nature of the observed changes in the SPS.

Table 4.1 summarizes the results showing that the gaze trajectories of test subjects with Parkinson's disease deviate notably from those of healthy controls. The inter-subject variance between healthy individuals is much less than that between healthy individuals and Parkinson patients, despite the age spread between the healthy subjects.

Figure 4.6 indicates that the parameters, and thereby also the dynamics, of the SPS in individuals with Parkinson's disease differ from those in healthy individuals.

The parameters of the SPS of the healthy controls vary depending on age, as can be seen in Fig. 4.6. However, it is also made apparent that deviations in the parameters due to Parkinson's disease are larger than the deviations due to age. The parameters of healthy and Parkinson subjects lie in different parameter domains even when the compared subjects are of similar age.

The results presented above indicate that there indeed are profound differences between the smooth pursuit eye movements of the considered individuals with Parkinson's disease and the healthy individuals. Although the number of participating test subjects in this study is small, and although no effort is made herein to explain the medical reason for the deviating parameters and gaze trajectories in Parkinson subjects, these results are still indicative of the potential of the presented methods and encourage future research to further investigate the methods as tools for diagnosing or staging of Parkinson's disease.

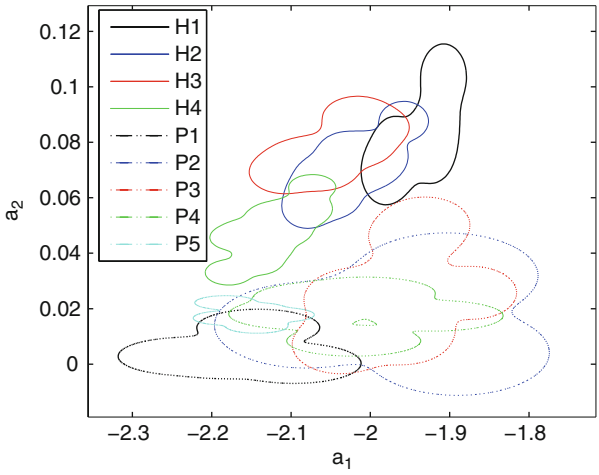


Fig. 4.6 The outlier regions of the estimated parameter distributions for different test subjects. *Dashed lines* are associated with test subjects with Parkinson's disease and *solid lines* with healthy controls (Reprinted with permission from [10]. Copyright 2013, AIP Publishing LLC)

Acknowledgements This chapter is in part financed by Advanced Grant 247035 from European Research Council entitled "*Systems and Signals Tools for Estimation and Analysis of Mathematical Models in Endocrinology and Neurology*".

References

1. G. Avanzini, F. Girotti, T. Carazeni, R. Spreafico, Oculomotor disorders in Huntington's chorea. J. Neurol. Neurosurg. Psychiatry **42**, 581–589 (1979)
2. R. Bednarik, T. Kinnunen, A. Mihaila, P. Fränti, Eye-Movements as a Biometric, in *Image Analysis*, ed. by H. Kalviainen, J. Parkkinen, A. Kaarna. Lecture Notes in Computer Science, vol. 3540 (Springer, Berlin/Heidelberg, 2005), pp. 780–789
3. R. Dodge, Five types of eye movements in the horizontal meridian plane of the field of regard. Am. J. Physiol. **8**, 307–329 (1903)
4. J.R. Fienup, Phase retrieval algorithms: a comparison. Appl. Opt. **21**, 2758–2769 (1982)
5. J.M. Gibson, R. Pimlott, C. Kennard, Ocular motor and manual tracking in Parkinson's disease and the effect of treatment. J. Neurol. **50**, 853–860 (1987)
6. H. He, J. Li, P. Stoica, *Waveform Design for Active Sensing Systems – A Computational Approach* (Camebridge University Press, New York, 2011)
7. D. Jansson, A. Medvedev, Dynamic smooth pursuit gain estimation from eye-tracking data, in *IEEE Conference on Decision and Control*, Orlando (2011)
8. D. Jansson, A. Medvedev, Visual stimulus design in parameter estimation of the human smooth pursuit system from eye-tracking data, in *IEEE American Control Conference*, Washington D.C. (2013)
9. D. Jansson, A. Medvedev, H.W. Axelson, Mathematical modeling and grey-box identification of the human smooth pursuit mechanism, in *IEEE Multi-conference on Systems and Control*, Yokohama, (2010)
10. D. Jansson, A. Medvedev, H.W. Axelson, D. Nyholm, Stochastic anomaly detection in eye-tracking data for quantification of motor symptoms in Parkinson's disease, in *International Symposium on Computational Models for Life Sciences*, Sydney, vol. 1559, November 2013, pp. 98–107
11. D. Jansson, O. Rosén, A. Medvedev, Non-parametric analysis of eye-tracking data by anomaly detection, in *IEEE European Control Conference*, Zürich (2013)
12. E.R. Kandel, J.H. Schwartz, T.M. Jessell, *Principles of Neural Science* (McGraw Hill, New York, 2000)
13. P. Kasprowski, Eye movements in biometrics, in *Biometric Authentication*, ed. by D. Maltoni, A.K. Jain. Lecture Notes in Computer Science 3087 (Springer, Berlin/Heidelberg, 2004), pp. 248–258
14. N. Kathmann, A. Hochrein, R. Uwer, B. Bondy, Deficits in gain of smooth pursuit eye movements in Schizophrenia and affective disorder patients and their unaffected relatives. Am. J. Psychiatry **160**, 696–702 (2003)
15. T.H. Koornwinder, R. Wong, R. Koekoek, R. Swarttouw, Orthogonal polynomials, in *NIST Handbook of Mathematical Functions* (Camebridge University Press, Cambridge/New York, 2010). ISBN 978-0521192255
16. S. Marino, E. Sessam, G. Di Lorenzo, P. Lanzafame, G. Scullica, A. Bramanti, F. La Rosa, G. Iannizzotto, P. Bramanti, P. Di Bella, Quantitative analysis of pursuit ocular movements in Parkinson's disease by using a video-based eye-tracking system. Eur. Neurol. **58**, 193–197 (2007)
17. C.H. Meyer, A.G. Lasker, D.A. Robinson, The upper limit of human smooth pursuit velocity. Vis. Res. **25**, 561–563 (1985)
18. T. Nakamura, R. Kanayama, R. Sano, M. Ohki, Y. Kimura, M. Aoyagi, Y. Koike, Quantitative analysis of ocular movements in Parkinson's disease. Acta Oto-Iaryngologica **111**, 559–562 (1991)

19. W.T. Newsome, R.H. Wurtz, M.R. Dürsteler, A. Mikami, Deficits in visual motion processing following ibotenic acid lesions of the middle temporal visual area of macaque monkey. J. Neurosci. **5**, 825–840 (1985)
20. C. Ramaker, J. Marinus, A.M. Stiggelbout, B.J. van Hilten, Systematic evaluation of rating scales for impairment and disability in Parkinson's disease. Mov. Disord. **17**(5), 867–876 (2002)
21. C. Rashbass, The relationship between saccadic and smooth tracking eye movements. J. Physiol. **159**, 326–338 (1961)
22. S.C. Schwartz, Estimation of probability density by an orthogonal series. Ann. Math. Stat. **38**, 1261–1265 (1967)
23. M. Tarter, R. Kronmal, On multivariate density estimates based on orthogonal expansions. Ann. Math. Stat. **41**, 718–722 (1970)
24. J.R. Thompson, P.R.A. Tapia, *Non-parametric Function Estimation, Modeling & Simulation*. Misc. Bks. (Society for Industrial and Applied Mathematics, Philadelphia, 1990)
25. J.A. Tropp, I.S. Dhillon, R.W. Heath, T. Strohmer, Designing structured tight frames via an alternating projection method. IEEE Trans. Inf. Theory **51**, 188–209 (2005)
26. O.B. White, J.A. Saint-Cyr, R.D. Tomlinson, J.A. Sharpe, Ocular motor deficits in Parkinson's disease, II. Control of the saccadic and smoothp ursuit systems. Oxf. J. Med. Brain **106**, 571–587 (1983)
27. T. Wigren, MATLAB software for recursive identification of Wiener systems. Systems and Control, Department of Information Technology, Uppsala University (2007)

Chapter 5
Identification of the Reichardt Elementary Motion Detector Model

Egi Hidayat, Alexander Medvedev, and Karin Nordström

Abstract The classical Hassenstein-Reichardt mathematical elementary motion detector (EMD) model is treated analytically. The EMD is stimulated with drifting sinusoidal gratings, which are often used in motion vision research, thus enabling direct comparison with neural responses from motion-sensitive neurones in the fly brain. When sinusoidal gratings are displayed on a cathode ray tube monitor, they are modulated by the refresh rate of the monitor. This generates a pulsatile signature of the visual stimulus, which is also seen in the neural response. Such pulsatile signals make a Laguerre domain identification method for estimating the parameters of a single EMD suitable, allowing estimation of both finite and infinite-dimensional dynamics. To model the response of motion-sensitive neurones with large receptive fields, a pool of spatially distributed EMDs is considered, with the weights of the contributing EMDs fitted to the neural data by a sparse estimation method. Such an EMD-array is more reliably estimated by stimulating with multiple sinusoidal gratings, since these provide higher spatial excitation than a single sinusoidal grating. Consequently, a way of designing the visual stimuli for a certain order of spatial resolution is suggested.

Keywords System identification • Insect vision • Motion detector • Spatial excitation • Visual stimuli

5.1 Background

Biological visual systems are generally accepted to compute local motion via so-called elementary motion detectors (EMDs; for recent review, see [2]). In the fly optic ganglia, lobula plate tangential cells (LPTCs) are believed to spatially pool

E. Hidayat (✉) • A. Medvedev
Department of Information Technology, Uppsala University, Uppsala, Sweden
e-mail: egi.hidayat@it.uu.se; alexander.medvedev@it.uu.se

K. Nordström
Department of Neuroscience, Uppsala University, Uppsala, Sweden
e-mail: karin.nordstrom@neuro.uu.se

C. Sun et al. (eds.), *Signal and Image Analysis for Biomedical and Life Sciences*,
Advances in Experimental Medicine and Biology 823,
DOI 10.1007/978-3-319-10984-8_5

the output from many EMDs (see e.g., [5]). The physiology of LPTCs, and the behavioral output of insects, closely match the predictions of the EMD [1]. Note, however, that the biological evidence does not come from recordings of individual EMDs, but indirectly from the spatially pooled neural or behavioural output.

Even though insect visual systems have been studied for decades, the mathematical tools for parameter estimation of the EMD model based on the solid principles of system identification are still lacking. To the authors' knowledge, system identification has rarely been applied to insect vision before. Black-box (quadratic) Volterra models with white input have been studied in e.g., [11, 13]. However, the relation of these models to the notion of EMD is not clear as the identification result is obtained in terms of Volterra kernels. Besides, the necessity of white noise stimuli is excessive and restrictive. Note that this chapter does not advocate the use of the classical Reichardt EMD model over other motion detection constructs, but simply provides mathematical tools for system identification of the former, as well as addresses the design of adequate visual stimuli. The developed methods are generalizable to more advanced EMD models such as the switched ones.

The main contributions of this chapter are as follows. The classical EMD model is revisited and a general closed-form expression for its output is derived. It is shown that in the case of non-symmetrical EMD dynamics and single harmonic input, the Fourier spectrum of the output contains both a single and double frequency of the input. In a symmetrical EMD, the double-frequency terms are cancelled. For an EMD identification approach based on Laguerre domain representation of the involved signals, it is demonstrated that both finite and infinite-dimensional dynamics of the linear block of the EMD model can be estimated from a pair of input-output pulses. Spatial excitation properties of multiple sinusoidal gratings are studied and a way of designing such for a given excitation order is provided. The latter result enables unambiguous estimation of the EMD weights in a layer model.

5.2 Mathematical Model of EMD

The mathematical model of EMD [7] can be summarized as follows:

$$v^+(t) = \int_0^t w^+(t-\theta)u^+(\theta)\,\mathrm{d}\theta, \quad v^-(t) = \int_0^t w^-(t-\theta)u^-(\theta)\,\mathrm{d}\theta,$$
$$y(t) = v^+u^- - v^-u^+, \tag{5.1}$$

where u^+ and u^- are the scalar inputs of the EMD, y is the output, and w^+ and w^- are the impulse responses of the low-pass filters in the input channels. Both w^+ and w^- are assumed here to be finite-dimensional, to simplify calculations.

Introduce two vectors: $v^T = \begin{bmatrix} v^+ & v^- \end{bmatrix}$; $u^T = \begin{bmatrix} u^+ & u^- \end{bmatrix}$ and the (unitary) skew-symmetric matrix

$$Q = \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}.$$

Then the EMD output can be represented as $y = v^T Q u$, which expression is a skew-symmetric bilinear form with the Laplace transform

$$Y(s) = \mathscr{L}\{y(t)\} = \frac{1}{2\pi j} \int_{-j\infty}^{+j\infty} V^T(p) Q U(s-p)\, \mathrm{d}p.$$

Here and further on, the capital letters represent, as usual, the Laplace transforms of the corresponding time-domain signals. Denoting $w(\cdot) = \operatorname{diag}\left[w^+(\cdot), w^-(\cdot)\right]$, the filtered input of the EMD in Laplace domain reads $V(s) = W(s)U(s)$.

Introduce $\bar{w}(\cdot) = Q^T w(\cdot)$, and let the integrand in the expression above vanish at an infinite arc in the left half-side of the complex plane. This assumption is justified by the experimentally observed overall low-pass dynamics of the EMD. Then, it follows that

$$Y(s) = \frac{1}{2\pi j} \oint_{\Omega} U^T(p) \bar{W}^T(p) U(s-p)\, \mathrm{d}p, \tag{5.2}$$

where Ω includes all the singularities of the integrand inside the left half-side of the complex plane. Notice that the contour integral has to be calculated over the singularities of the linear part of the EMD, as well as the singularities of the Laplace transform of the EMD input and its s-shifted version.

Making use of the assumption that $w(t)$ is the impulse response of a finite-dimensional linear system, write

$$w(t) = \operatorname{diag}\left[C^+ \exp(A^+ t) B^+,\, C^- \exp(A^- t) B^-\right],$$

where (A^+, B^+, C^+) and (A^-, B^-, C^-) are minimal realizations of the transfer functions $C^+(pI - A^+)^{-1}B^+$ and $C^-(pI - A^-)^{-1}B^-$, correspondingly, with $A^+ \in \mathbb{R}^{n^+ \times n^+}$ and $A^- \in \mathbb{R}^{n^- \times n^-}$.

To obtain a time domain expression of (5.2), the contour integral is solved by applying Cauchy's residue theorem. Rewrite (5.2) as $Y(s) = Y^+(s) - Y^-(s)$ with

$$Y^+(s) = \frac{1}{2\pi j} C^+ \oint_{\Omega} (pI - A^+)^{-1} U^+(p) U^-(s-p)\, \mathrm{d}p\; B^+, \tag{5.3}$$

$$Y^-(s) = \frac{1}{2\pi j} C^- \oint_{\Omega} (pI - A^-)^{-1} U^-(p) U^+(s-p)\, \mathrm{d}p\; B^-. \tag{5.4}$$

To evaluate the integrals above, the signal shape of the input has to be known. In biological motion vision research, drifting sinusoidal gratings are often used as stimuli. The results below will be specialized to this practically important case.

5.2.1 Single Frequency Sinusoidal Signal

The spatial sampling of the EMD manifests itself in the form of a phase shift between the two inputs. In the case of one sinusoidal signal of frequency ω passing from one input to another, the movement results in a phase shift δ whose value depends on the velocity of the movement and the spatial separation of the involved photoreceptors, i.e., $u^+(t) = \sin(\omega t)$, $u^-(t) = \sin(\omega t + \delta)$. The output signal for each block is given as follows:

$$\begin{aligned} y^+(t) = & C^+(\omega^2 I + (A^+)^2)^{-1}\big(\omega \sin(\omega t + \delta)\exp(A^+ t) \\ & - \tfrac{1}{2}\omega\left(\sin(2\omega t + \delta) + \sin(\delta)\right) I + \tfrac{1}{2}A^+\left(\cos(2\omega t + \delta) - \cos(\delta)\right)\big) B^+, \end{aligned} \tag{5.5}$$

$$\begin{aligned} y^-(t) = & C^-(\omega^2 I + (A^-)^2)^{-1}\big((A^- \sin(\delta) + \omega\cos(\delta) I)\sin(\omega t)\exp(A^- t) \\ & - \tfrac{1}{2}\omega\left(\sin(2\omega t + \delta) - \sin(\delta)\right) I + \tfrac{1}{2}A^-\left(\cos(2\omega t + \delta) - \cos(\delta)\right)\big) B^-. \end{aligned} \tag{5.6}$$

By inspection of the expressions above, it becomes clear that the resulting output is composed by two frequency components. The first component has the same frequency as the input signal. This component appears only in the transient and will eventually decay due to the converging exponential factor. The second component has the double frequency compared to that of the input signal. This component constitutes the steady-state solution of the EMD model.

5.2.1.1 Symmetrical and Non-symmetrical EMD Model

The EMD model considered in biological motion vision is typically symmetrical, i.e., $w^+(\cdot)$ and $w^-(\cdot)$ in (5.1) are identical [3]. In this situation, the EMD output in (5.5) and (5.6) is reduced to

$$y(t) = C(\omega^2 I + A^2)^{-1}\big(\big(\omega\cos(\omega t) I - A\sin(\omega t)\big)\exp(At) - \omega\big) B\sin(\delta),$$

which agrees with the one given in [3]. In the symmetrical structure, the terms related to the double-frequency cancel each other, thus leaving the steady-state solution independent of the time variable. For a non-symmetrical structure, the double-frequency component would persist in the steady-state solution, though this is often disregarded [14]. Notably, early publications, e.g., [13], observed, without a mathematical derivation, that the double-frequency harmonics have significant impact on the output signal, in line with what has been shown in the previous section.

5.2.2 *EMD Response to a $\mathbb{L}_2$ Pulse*

Sinusoidal gratings are often used in motion vision research. These can be displayed to the animal using a range of different display systems, such as DLP projectors, LCD screens, LED arenas or CRT monitors. When using a display with a limited refresh rate, such as a CRT monitor, the sinusoidal grating will be displayed as a change of luminance on the screen, and simultaneously modulated by the screen refresh rate. Figure 5.1a shows the resulting luminance change as a function of time at one pixel of the screen (measured by a photodiode transducer) when displaying a 20 Hz sinusoidal grating on a 160 Hz CRT monitor. The pulsatile high frequency component that appears throughout the photodiode measurement corresponds to the refresh rate of the monitor.

Each input pulse of the visual stimulus sequence evokes a corresponding pulse in the output of the EMD. Thus, both pulses can be thought of as signals that belong to $\mathbb{L}_2$, i.e., functions integrable with square. In this class of signals, identification in the Laguerre domain, cf. [8], is suitable and therefore considered in this chapter. Furthermore, biological systems often generate highly variable responses. Such variability can be seen in the neural data. Despite being excited by four cycles of

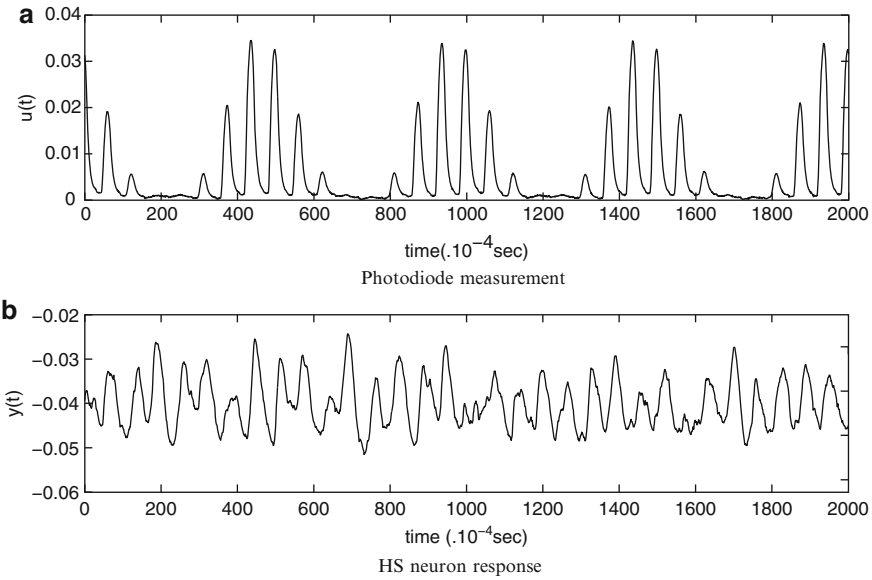


Fig. 5.1 Input and output signal forms. (**a**) Photodiode measurement of sinusoidal grating (20 Hz, wavelength 50 pixels) displayed on a 160 Hz CRT monitor. (**b**) Intracellular response of a motion sensitive neurone in the fly brain to the stimulus in panel (**a**). Note that the screen refresh rate at 160 Hz is much more apparent in the neural data, than the sinusoidal motion 20 Hz (Reprinted with permission from [10]. Copyright 2013, AIP Publishing LLC)

identical input (Fig. 5.1a), the neural response is highly variable (Fig. 5.1b). The use of Laguerre domain identification helps to circumvent the effects of such biological variability on the accuracy of the parameter estimates by utilizing one pair of pulses at a time.

5.3 Identification Approach

The data shown in Fig. 5.1b, are recorded from an LPTC in the fly brain. As mentioned above, these tangential neurons are believed to spatially pool the output from many local EMDs. The identification process of EMDs from neural recordings considered below comprises two stages. The first stage handles the estimation of the dynamics of the single EMD model, while the second stage estimates the interconnection structure between the contributing individual EMDs in the layer that are believed to contribute to the response of an LPTC.

5.3.1 Identification of a Single EMD

The method for identification of the EMD model by means of Laguerre functions was introduced in [9]. It utilizes the representation of input and output signals in terms of the Laguerre spectra, and also the representation of the EMD dynamics in the Laguerre domain.

The Laguerre spectrum of a signal is obtained by evaluating the inner functional product in $\mathbb{L}_2$ between the signal itself and the Laguerre functions. The continuous Laguerre function of order k is given by

$$\ell_k(t) = \sqrt{2\alpha}\mathrm{e}^{-\alpha t} \sum_{n=0}^{k} \binom{k}{n} \frac{(-2\alpha t)^n}{n!},$$

where $\alpha > 0$ is the Laguerre parameter.

It is shown in [9] that the original nonlinear EMD model can be identified by solving a linear problem that expresses the relationship between the Markov parameters of the linear blocks in Laguerre domain and the Laguerre spectrum of the output signal

$$Y_q = \Pi_q \Gamma_k, \quad Y_q = \left[y_0 \ \ldots \ y_q \right]^T, \tag{5.7}$$

where Γ_k is a vector of the Markov parameters of the linear blocks in Laguerre domain, and y_i are the Laguerre coefficients of y for $i \in \{0, q\}$. The matrix Π_q is given by $\Pi_q = \left[\Omega_0 \ \ldots \ \Omega_q \right]^T$, where $\Omega_i = \sqrt{2\alpha} \left[(U_k^+)^T \ (U_k^-)^T \right] \Xi_i$, $i = \{0, q\}$.

The terms U_k^+ and U_k^- define the Laguerre spectra of the input signals. The matrix Ξ_q is calculated as a linear combination of the input signals spectra

$$\Xi_q = \begin{bmatrix} -\Upsilon_q & \Psi_k^- \\ \Upsilon_q & \Psi_k^+ \end{bmatrix},$$

where Ψ_k^+ and Ψ_k^- are the lower triangular Toeplitz matrices consisting of the Laguerre coefficients of the respective input signals and the matrix

$$\Upsilon_q = \begin{bmatrix} C_{00q} & \cdots & C_{0kq} \\ \vdots & & \vdots \\ C_{k0q} & \cdots & C_{kkq} \end{bmatrix},$$

comprises the constants given by

$$C_{mnq} = \frac{2}{3} \sum_{r=0}^{m} \sum_{s=0}^{n} \sum_{\gamma=0}^{q} \left(-\frac{2}{3}\right)^{r+s+\gamma} \binom{m}{r} \binom{n}{s} \binom{q}{\gamma} \binom{r+s+\gamma}{r,s,\gamma}, \tag{5.8}$$

$$\text{with } \binom{r+s+\gamma}{r,s,\gamma} = \frac{(r+s+\gamma)!}{r!s!\gamma!}.$$

Notice that Υ_q can be evaluated once and for all since it is independent both of the input-output data and the dynamics of the EMD. For more details on the identification approach, see [9].

5.3.1.1 Pure Time-Delay Model

The linear blocks in the EMD model are often considered to be pure time-delays [1]. This assumption can be enforced in system identification via a finite-dimensional approximation of the time-delay operator by, e.g., Padé or Laguerre series [4, 6, 12]. The approximation approach imposes a certain structure of the dynamics of the linear part and results in a grey-box identification setting. When finite-dimensional dynamics are identified, an estimate of the time delay can be evaluated from the relationships between the approximations terms. A direct approach to time delay dynamics identification is suggested here instead.

Let the linear block $w(\cdot)$ represent the pure time-delay operator, i.e., $v(t) = u(t - t_d)$, with a time-delay t_d. The output of the linear block in Laguerre domain can then be written as [4] $V_k^o = \Phi_k^o \Theta_k$, where the regression matrix Φ_k^o is given by

$$\Phi_k^o = \begin{bmatrix} \varphi_0^o & \dots & \varphi_k^o \end{bmatrix}^T, \quad \varphi_{k,l+1}^o = \begin{cases} u_k^o & l = 0, \\ \frac{(-2)^l}{l!(l-1)!} \sum_{i=0}^{k-l} \frac{(k-i-1)!}{(k-i-l)!} u_i & k \geq l > 0, \\ 0 & l > k, \end{cases}$$

and the parameter vector Θ_k is defined as

$$\Theta_k = \begin{bmatrix} \theta_0 \ \dots \ \theta_k \end{bmatrix}^T, \quad \theta_i = (\alpha t_d)^{i-1} \mathrm{e}^{-\alpha t_d}, i \in \{0, k\}.$$

The superscript $(\cdot)^o$ means either $(\cdot)^-$ or $(\cdot)^+$.

Since the relation between the outputs of the linear blocks and the EMD output still holds, only a simple modification in the identification approach is required to identify an EMD with pure-time delay linear blocks. Replacing Ψ_k^o and Γ_k with Φ_k^o and Θ_k will give the relation between the time-delay operator and the Laguerre coefficients of the EMD output as

$$y_q = \sqrt{2\alpha} \begin{bmatrix} (U_k^+)^T & (U_k^-)^T \end{bmatrix} \tilde{\Xi}_q \Theta_k, \text{ where } \tilde{\Xi}_q = \begin{bmatrix} -\Upsilon_q & \Phi_k^- \\ \Upsilon_q & \Phi_k^+ \end{bmatrix}.$$

Hence, the linear relationship between the EMD output Laguerre spectrum and the delay parameter vector Θ_k can be derived in similar manner as in (5.7).

5.3.2 *Identification of a Layer of EMDs*

Despite recent advances in understanding motion detection in insect vision using genetic tools in *Drosophila* [2] it is still impossible to record the output of a single EMD. However, a signal that corresponds to the weighted sum of the outputs of many EMDs can be measured in fly LPTCs, such as in Fig. 5.1b. Therefore, systems of multiple EMDs subject to visual stimulation have to be considered in order to apply system identification to experimental data.

It is assumed that the EMDs are uniformly spatially distributed in a flat layer and all possess identical dynamics as in (5.1). Under this assumption, the EMDs are exposed to spatially identical visual input, albeit shifted in time.

Let $y(t, n)$ denote the output of the n-th EMD in the layer, then the measured signal of the EMD-layer output is therefore modeled as

$$y_N(t) = \sum_{j=1}^{N} g_j y(t, j) = \sum_{j=1}^{N} g_j y(t - \tau_j), \tag{5.9}$$

where $g_j, \tau_j, j = 1, \dots, N$ are unknown constants characterizing the contribution of an individual EMD. With sparse estimation in mind, the following optimization problem can be formulated to calculate the estimates of the EMD weights g_j and time delays τ_j:

$$[\hat{g}, \hat{\tau}] = \arg\min_{g,\tau} \sum_{i=1}^{M} \left(y(t_i) - \sum_{j=1}^{N} g_j \hat{y}(t_i - \tau_j) \right)^2, \tag{5.10}$$

$$\text{s.t.} \sum_{j=1}^{N} |g_j|_1 < \varepsilon.$$

The constraint in (5.10) penalizes over-parametrized signal models and helps to keep the number of participating EMDs (i.e., N) as low as possible. Optimization problems as in (5.10) are considered to be difficult to solve. One possible approach is to compute a suboptimal solution to the problem by assuming that $\tau_j, j = 1, \ldots, N$ are commensurate or belong to a given grid set. Lowering the threshold value for the weights results in a reduction of the number of the contributing EMDs.

5.3.2.1 Identifiability Properties for EMD-Layer Estimation

The expression for the output of EMD-layer in (5.9) is derived based on the assumption that there are N spatially distributed contributing EMDs in the layer. Clearly, two EMDs receiving the same input cannot be distinguished among since they possess identical dynamics. In order to obtain a unique solution for N unknowns, at least N linearly independent equations are required. The excitation order of the stimulus in terms of spatial distribution plays an important role in preventing the linear dependency between the individual EMD outputs. Here, a study of the spatial excitation order of a sinusoidal grating is presented.

Single Frequency Sinusoidal Grating

The sinusoidal grating stimuli exciting the two input channels of the n-th EMD in a layer are described as follows:

$$u^+(t,n) = c_0 + c_1 \sin(\omega t + \phi + \tilde{n}\delta), \quad u^-(t,n) = c_0 + c_1 \sin(\omega t + \phi + n\delta),$$

with $\tilde{n} = n - 1$. The stimuli are thus parametrized in terms of the frequency ω and the phase shift ϕ. The weights c_0 and c_1 represent the mean luminance of the pattern and the pattern contrast of the grating used in fly experiments. The constant δ is the phase difference between the two input channels resulting from the temporal delay between the channels τ that is described by

$$\delta = -\omega\tau.$$

Without loss of generality, an EMD with a first-order linear block is treated. For the symmetrical EMD model with the linear block transfer function

$$W(s) = \frac{K}{s + a},$$

the EMD output is given by

$$y(t,n) = y_t(t,n) + y_s(t,n),$$

where $y_t(t,n)$ and $y_s(t,n)$ represent the transient and the steady-state response, respectively. It can be shown by a straightforward calculation that the transient response is

$$\begin{aligned} y_t(t,n) = \frac{2c_1 K \sin\left(\frac{\delta}{2}\right)}{a\sqrt{a^2+\omega^2}} \mathrm{e}^{-at} \Bigg[& c_1 a \cos\left(\frac{\delta}{2}\right) \sin\left(\omega t + \tan^{-1}\left(\frac{\omega}{a}\right)\right) \\ & + c_0 \sin(n\delta)\left(\sqrt{a^2+\omega^2}\sin\left(\omega t + \phi - \frac{\delta}{2}\right) + a\cos\left(\phi - \frac{\delta}{2} + \tan^{-1}\left(\frac{a}{\omega}\right)\right)\right) \\ & + c_0 \cos(n\delta)\left(-\sqrt{a^2+\omega^2}\cos\left(\omega t + \phi - \frac{\delta}{2}\right) + a\sin\left(\phi - \frac{\delta}{2} + \tan^{-1}\left(\frac{a}{\omega}\right)\right)\right)\Bigg], \end{aligned} \tag{5.11}$$

while the steady-state response is given by

$$\begin{aligned} y_s(t,n) = \frac{2c_1 K \sin\left(\frac{\delta}{2}\right)}{a\sqrt{a^2+\omega^2}} \Bigg[& -\frac{c_1 a\omega \cos\left(\frac{\delta}{2}\right)}{\sqrt{a^2+\omega^2}} - c_0 \sin(n\delta)\sin\left(\omega t + \phi - \frac{\delta}{2} + \tan^{-1}\left(\frac{a}{\omega}\right)\right) \\ & + c_0 \cos(n\delta)\cos\left(\omega t + \phi - \frac{\delta}{2} + \tan^{-1}\left(\frac{a}{\omega}\right)\right)\Bigg]. \end{aligned} \tag{5.12}$$

From (5.11) and (5.12), it is seen that the n-th EMD's response to a sinusoidal grating comprises three linearly independent elements with respect to n. Since the decay rate of the transient EMD response is relatively high due to the fast linear dynamics, this identifiability analysis focuses on the steady state response.

To simplify the computations, define $\tilde{y}_s(t,n)$ as a normalized version of (5.12) such that

$$\tilde{y}_s(t,n) = \tilde{c} - \sin(\omega t + \tilde{\phi})\sin(n\delta) + \cos(\omega t + \tilde{\phi})\cos(n\delta), \tag{5.13}$$

where

$$\begin{aligned} \tilde{y}_s(t,n) &= y_s(t,n)\frac{a\sqrt{a^2+\omega^2}}{2c_0 c_1 K\omega \sin(\frac{\delta}{2})}, \\ \tilde{c} &= -\frac{ac_1\cos(\frac{\delta}{2})}{c_0\sqrt{a^2+\omega^2}}, \\ \tilde{\phi} &= \phi - \frac{\delta}{2} + \tan^{-1}\left(\frac{a}{\omega}\right). \end{aligned}$$

From (5.9) and (5.13), the normalized EMD-layer steady-state output is

$$\tilde{y}_{N,s}(t) = \tilde{c}\sum_{n=1}^{N} g_n - \sin(\omega t + \tilde{\phi})\sum_{n=1}^{N} g_n \sin(n\delta) + \cos(\omega t + \tilde{\phi})\sum_{n=1}^{N} g_n \cos(n\delta).$$

The expression above can be represented in a matrix form as

$$Y = \Xi \boldsymbol{\gamma},$$
$$\Xi = \tilde{c} J - \tilde{f_s} \otimes z_s + \tilde{f_c} \otimes z_c,$$

where J is an $(M+1) \times N$ matrix of unit elements, $\otimes$ the tensor product of vectors, and

$$\tilde{f_s} = \left[\sin(\tilde{\phi}) \ \sin(\omega T + \tilde{\phi}) \ \ldots \ \sin(M\omega T + \tilde{\phi})\right]^T$$
$$\tilde{f_c} = \left[\cos(\tilde{\phi}) \ \cos(\omega T + \tilde{\phi}) \ \ldots \ \cos(M\omega T + \tilde{\phi})\right]^T$$
$$z_s = \left[\sin(\delta) \ \sin(2\delta) \ \ldots \ \sin(N\delta)\right]^T$$
$$z_c = \left[\cos(\delta) \ \cos(2\delta) \ \ldots \ \cos(N\delta)\right]^T.$$

The vector Y stands for the EMD-layer steady-state output vector and the vector $\boldsymbol{\gamma}$ is comprised of the weights of the participating EMDs

$$Y = \left[\tilde{y}_{N,s}(0) \ \tilde{y}_{N,s}(T) \ \ldots \ \tilde{y}_{N,s}(MT)\right]^T,$$
$$\boldsymbol{\gamma} = \left[g_1 \ g_2 \ \ldots \ g_N\right]^T.$$

The rank of Ξ determines the largest number of EMD weights that can be uniquely estimated from Y. As it always holds that rank $\Xi \leq 3$, one concludes that only three EMDs can be uniquely distinguished among by means of a one-frequency sinusoidal grating, i.e., $n \leq 3$.

Sum of Sinusoidal Gratings

It has been shown in Sect. 5.3.2.1 that a single-frequency sinusoidal grating stimulus, such typically used in motion vision experiments is not sufficient for identifying a large number of participating EMDs in a layer. This is important since hundreds of EMDs are expected to contribute to responses recorded in LPTCs and behaviour. To enable an unambiguous estimation of an EMD layer, the stimuli therefore have to possess a high spatial excitation order. One feasible solution is to use a stimulus composed of the sum of several sinusoidal gratings. This is well in line with Fourier series as a means of representing bounded periodical signals.

For r sinusoidal gratings, the input signals to the n-th EMD are defined as follows:

$$u^+(t,n) = c_0 + \sum_{i=1}^{r} c_i \sin(\omega_i t + \phi_i + \tilde{n}\delta_i),$$
$$u^-(t,n) = c_0 + \sum_{i=1}^{r} c_i \sin(\omega_i t + \phi_i + n\delta_i)$$

To avoid unnecessary cumbersome expressions, the derivation below is carried out for the case of $r = 2$, but it can in a straightforward manner be generalized for an arbitrary r by expanding the involved sums.

Similarly to the single-frequency case, the EMD response to a sum of sinusoidal gratings comprises a transient and a steady-state component. The transient response can be written as

$$
\begin{aligned}
y_t(t,n) = {} & \frac{2\mathrm{e}^{-at}}{a} \sum_{i=1}^{2} \frac{c_i K \sin\left(\frac{\delta_i}{2}\right)}{\sqrt{a^2+\omega_i^2}} \Bigg[c_i a \cos\left(\frac{\delta_i}{2}\right) \sin\left(\omega_i t + \tan^{-1}\left(\frac{\omega_i}{a}\right)\right) \\
& + c_0 \sin(\tilde{n}\delta_i) \left(\sqrt{a^2+\omega_i^2} \sin\left(\omega_i t + \phi_i + \frac{\delta_i}{2}\right) + a \cos\left(\phi_i + \frac{\delta_i}{2} + \tan^{-1}\left(\frac{a}{\omega_i}\right)\right)\right) \\
& + c_0 \cos(\tilde{n}\delta_i) \left(-\sqrt{a^2+\omega_i^2} \cos\left(\omega_i t + \phi_i + \frac{\delta_i}{2}\right) + a \sin\left(\phi_i + \frac{\delta_i}{2} + \tan^{-1}\left(\frac{a}{\omega_i}\right)\right)\right) \Bigg] \\
& + \frac{c_1 c_2 K}{a^2+\omega_1^2} \mathrm{e}^{-at} \Big[\Big(\omega_1 \cos(\phi_1 + \tilde{n}\delta_1) - a \sin(\phi_1 + \tilde{n}\delta_1) \Big) \sin(\omega_2 t + \phi_2 + n\delta_2) \\
& - \Big(\omega_1 \cos(\phi_1 + n\delta_1) - a \sin(\phi_1 + n\delta_1) \Big) \sin(\omega_2 t + \phi_2 + \tilde{n}\delta_2) \Big] \\
& + \frac{c_1 c_2 K}{a^2+\omega_2^2} \mathrm{e}^{-at} \Big[\Big(\omega_2 \cos(\phi_2 + \tilde{n}\delta_2) - a \sin(\phi_2 + \tilde{n}\delta_2) \Big) \sin(\omega_1 t + \phi_1 + n\delta_1) \\
& - \Big(\omega_2 \cos(\phi_2 + n\delta_2) - a \sin(\phi_2 + n\delta_2) \Big) \sin(\omega_1 t + \phi_1 + \tilde{n}\delta_1) \Big],
\end{aligned}
$$

while the steady state response reads as

$$
\begin{aligned}
y_s(t,n) = {} & \sum_{i=1}^{2} \frac{2c_i K \sin\left(\frac{\delta_i}{2}\right)}{a\sqrt{a^2+\omega_i^2}} \Bigg[- \frac{c_i a \omega_i \cos\left(\frac{\delta_i}{2}\right)}{\sqrt{a^2+\omega_i^2}} \\
& + c_0 \cos\left(\omega t + \left(n - \frac{1}{2}\right)\delta_i + \phi_i + \tan^{-1}\left(\frac{a}{\omega_i}\right)\right) \Bigg] \\
& + \frac{c_1 c_2 K}{a^2+\omega_1^2} \Big[\Big(a \sin(\omega_1 t + \phi_1 + \tilde{n}\delta_1) - \omega_1 \cos(\omega_1 t + \phi_1 + \tilde{n}\delta_1) \Big) \sin(\omega_2 t + \phi_2 + n\delta_2) \\
& - \Big(a \sin(\omega_1 t + \phi_1 + n\delta_1) - \omega_1 \cos(\omega_1 t + \phi_1 + n\delta_1) \Big) \sin(\omega_2 t + \phi_2 + \tilde{n}\delta_2) \Big] \\
& + \frac{c_1 c_2 K}{a^2+\omega_2^2} \Big[\Big(a \sin(\omega_2 t + \phi_2 + \tilde{n}\delta_2) - \omega_2 \cos(\omega_2 t + \phi_2 + \tilde{n}\delta_2) \Big) \sin(\omega_1 t + \phi_1 + n\delta_1) \\
& - \Big(a \sin(\omega_2 t + \phi_2 + n\delta_2) - \omega_2 \cos(\omega_2 t + \phi_2 + n\delta_2) \Big) \sin(\omega_1 t + \phi_1 + \tilde{n}\delta_1) \Big].
\end{aligned}
\tag{5.14}
$$

Once again, neglecting the transient component that vanishes at a high convergence rate, the main focus of the analysis is on the sustained steady-state response of the EMD. Introduce the following notation

$$\omega_{i,j}^{+} = \omega_i + \omega_j, \quad \delta_{i,j}^{+} = \delta_i + \delta_j, \; \omega_{i,j}^{-} = \omega_i - \omega_j, \quad \delta_{i,j}^{-} = \delta_i - \omega_j,$$
$$\tilde{c} = -\sum_{i=1}^{r} \frac{c_i^2 \omega_i K \sin(\delta_i)}{a^2+\omega_i^2}, \qquad \tilde{c}_i = \frac{2c_0 c_i K \sin(\frac{\delta_i}{2})}{a\sqrt{a^2+\omega_i^2}}, \quad \tilde{\phi}_i = \phi_i - \frac{\delta_i}{2} + \tan^{-1}\left(\frac{a}{\omega_i}\right)$$
$$\tilde{c}_{i,j}^{+} = \frac{c_i c_j \omega_{i,j}^{-} K \sin\left(\frac{\delta_{i,j}^{-}}{2}\right)}{\sqrt{(a^2+\omega_i^2)(a^2+\omega_j^2)}}, \qquad \tilde{\phi}_{i,j}^{+} = \phi_i + \phi_j - \frac{\delta_{i,j}^{+}}{2} + \tan^{-1}\left(\frac{a^2-\omega_i\omega_j}{a\omega_{i,j}^{+}}\right)$$
$$\tilde{c}_{i,j}^{-} = -\frac{c_i c_j \omega_{i,j}^{+} K \sin\left(\frac{\delta_{i,j}^{+}}{2}\right)}{\sqrt{(a^2+\omega_i^2)(a^2+\omega_j^2)}}, \qquad \tilde{\phi}_{i,j}^{-} = \phi_i - \phi_j - \frac{\delta_{i,j}^{-}}{2} + \tan^{-1}\left(\frac{a^2+\omega_i\omega_j}{a\omega_{i,j}^{-}}\right).$$

With the definitions above, the expression in (5.14) can be represented in a more compact form as follows

$$y_s(t,n) = \tilde{c} + \sum_{i=1}^{r} \tilde{c}_i \cos(\omega_i t + n\delta_i + \tilde{\phi}_i)$$
$$+ \sum_{i=1}^{r-1} \sum_{j=i+1}^{r} \left(\tilde{c}_{i,j}^{+} \sin(\omega_{i,j}^{+} t + n\delta_{i,j}^{+} + \tilde{\phi}_{i,j}^{+}) + \tilde{c}_{i,j}^{-} \sin(\omega_{i,j}^{-} t + n\delta_{i,j}^{-} + \tilde{\phi}_{i,j}^{-}) \right), \tag{5.15}$$

with $r = 2$. Equation (5.15) reveals the general structure of the steady state response of the n-th EMD to the sum of r sinusoidal gratings. The spatial excitation order corresponds to the number of linearly independent functions of n in (5.15), which quantity is easily seen to be at most $2r^2 + 1$. By inspection of (5.15), it follows that the number of independent components of the output falls under the this bound whenever

$$\omega_{i,j}^{+}, \omega_{i,j}^{-} \in \{\omega_i, i = 1, \ldots, r\}.$$

5.3.2.2 Spatial Excitation of a Sum of Sinusoidal Gratings: An Example

In the study of the sum of sinusoidal gratings, the excitation problem is considered in two dimensions with the spatiotemporal signal as a function of time (the variable t) and space (the variable n). The transformation from the spatial distance to the phase shift of an individual sine wave determines the excitation order of the stimulus. Here, two cases are presented to illustrate the significance of the phase shift formulation.

- Stimulus Type 1: Constant time interval between two locations (Fig. 5.2, top panel). In this case each sine wave is characterized by a unique spatial wavelength. For instance, to propagate as far as 100 pixels within 1 s, a sine wave

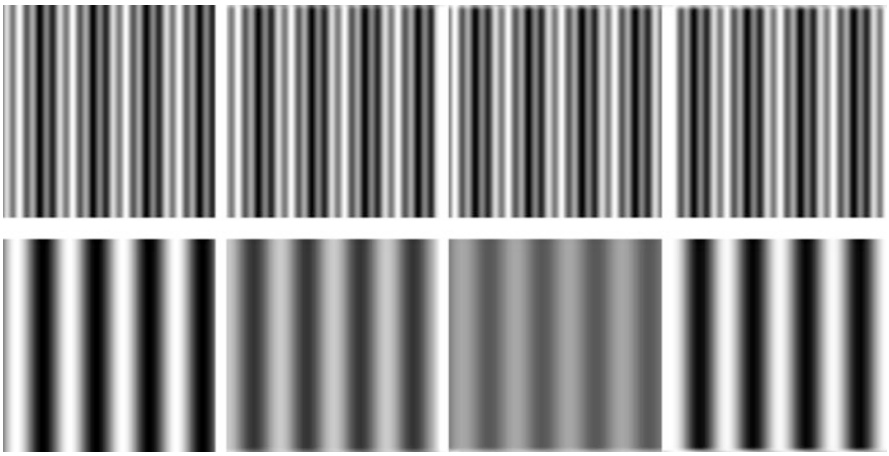


Fig. 5.2 Propagation of two sine wave grating stimulus. *Top row*: stimulus of Type 1 – constant time interval. *Bottom row*: stimulus of Type 2 – constant phase shift

with frequency of 10 Hz needs spatial wavelength of 10 pixels, while a sine wave with frequency of 20 Hz requires spatial wavelength of 5 pixels. For this stimulus type, the input to the second input channel of the n-th EMD is then given by

$$\begin{aligned} u^{-}(t,n) &= \sum_{i=1}^{r} \sin(\omega_i t + n\delta_i) \\ &= \sum_{i=1}^{r} \sin(\omega_i (t - n\tau)) \end{aligned}$$

which gives $\delta_i = -\omega_i \tau$.

- Stimulus Type 2: Constant spatial wavelength (Fig. 5.2, bottom panel). In this case each sine wave is characterized by the time interval elapsed while the wave travels between the two locations. For example, to propagate as far as 100 pixels with spatial wavelength of 10 pixels, a sine wave with frequency of 10 Hz will need 1 s, while a sine wave with frequency of 20 Hz will take 0.5 s. For this stimulus type, the input to the second channel on the n-th EMD is then described as

$$\begin{aligned} u^{-}(t,n) &= \sum_{i=1}^{r} \sin(\omega_i t + n\delta_i) \\ &= \sum_{i=1}^{r} \sin(2\pi f_i (t - n\tau_i)) \end{aligned}$$

$$= \sum_{i=1}^{r} \sin\left(2\pi f_i \left(t - \frac{nd}{\lambda f_i}\right)\right)$$
$$= \sum_{i=1}^{r} \sin\left(\omega_i t - 2\pi \frac{nd}{\lambda}\right),$$

which results in $\delta_i = -2\pi d/\lambda$.

Clearly, for a stimulus of Type 2, the choice of ω_i does not have any effect on δ_i. Using such a stimulus, the number of linearly independent components in (5.15) will be greatly reduced as, for any i, $\delta_i = \delta$ and $\delta_{\bar{i,j}} = 0$. Therefore, the output signal obeys

$$y_s(t,n) = \tilde{c} + \sum_{i=1}^{r} \tilde{c}_i \cos(\omega_i t + n\delta + \tilde{\phi}_i) + \sum_{i=1}^{r-1} \sum_{j=i+1}^{r} \tilde{c}_{\bar{i,j}} \sin(\omega_{\bar{i,j}} t + \tilde{\phi}_{\bar{i,j}}).$$

The above expression comprises three linearly independent functions at most, whose structure yields the same spatial excitation order as stimulating with a single sinusoidal grating.

5.3.2.3 Visualization

Graphical illustrations of the two stimulus types are provided here to support the theoretical discussion in the preceding section. The temporal frequencies of the sinusoidal gratings are selected to be 5 and 20 Hz and the patch size is set to 200×200 pixels. Figure 5.2 depicts the resulting 2D screenshots at four different times. The figures in the top row show that the Type 1 stimulus produces a continuous pattern moving in one direction at a constant velocity. The frequencies and weights of the sine waves create spatial variability in the pattern. The figures in the bottom row show the corresponding images generated with the stimulus of Type 2. In this type of stimulus, each sine wave drifts across the patch at its own unique velocity. In this case, the four screenshots show that the stimulus is visualized as a modulation of several sine waves.

In Fig. 5.3, the two stimulus types are compared by the local luminance change at three spatially separated locations (illustrated with different colours). The three locations are separated by six pixels in the direction of motion. The top plot shows that in the Type 1 stimulus, an identical luminance signal is simply delayed between the three locations. However, the bottom plot reveals that the Type 2 stimulus generates similar waveforms at different locations. Therefore, the EMDs in the layer receive similar excitation and cannot be discriminated between. As an immediate result, the layer structure represented by the weights of the contributed EMDs cannot be uniquely estimated.

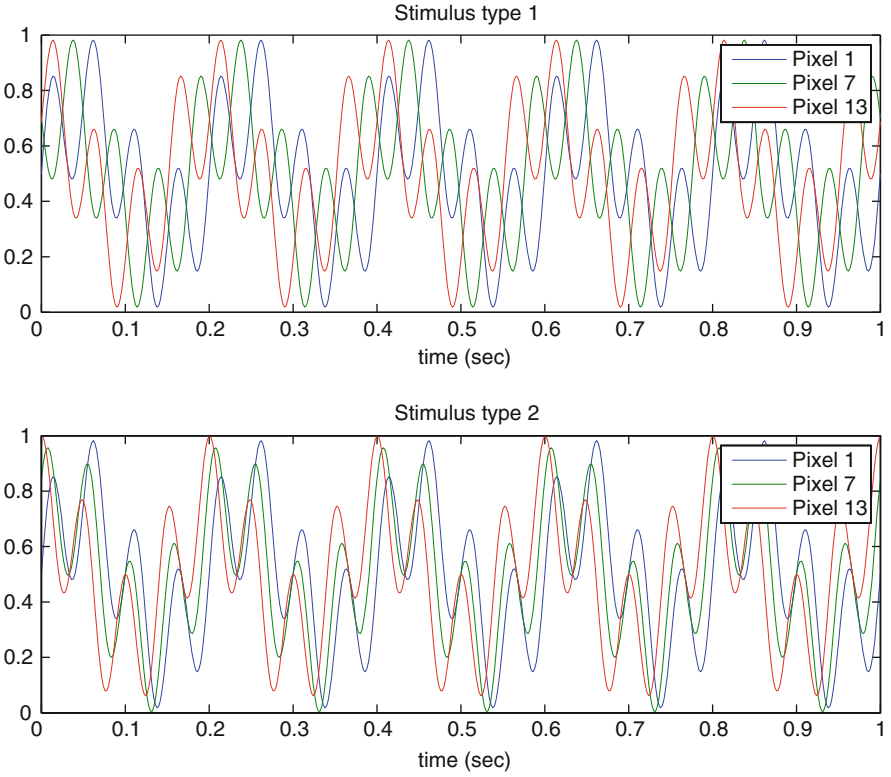


Fig. 5.3 Input signals at three different locations using two different formulations for visual stimulation with a sum of sine waves

To test the efficacy of the above described types of visual stimuli, a layer of 50 individual EMDs with randomly selected weight values is simulated. The performance of the layer estimation by solving the sparse optimization problem in (5.10) is evaluated with stimuli of the two types. As seen in the top plot in Fig. 5.4, the weight estimation in Type 1 improves substantially with the number of sine waves in the stimulus. The spatial excitation order of the signal with $r = 5$ is $2r^2 + 1 = 51$, which is sufficient for unambiguous estimation of the weights of all the 50 EMDs constituting the layer.

The bottom plot in Fig. 5.4 demonstrates that with the stimulus of Type 2, the weight estimates do not improve with the number of sine waves. This simulation confirms the statement above about the weak spatial excitation ability of the stimuli of Type 2.

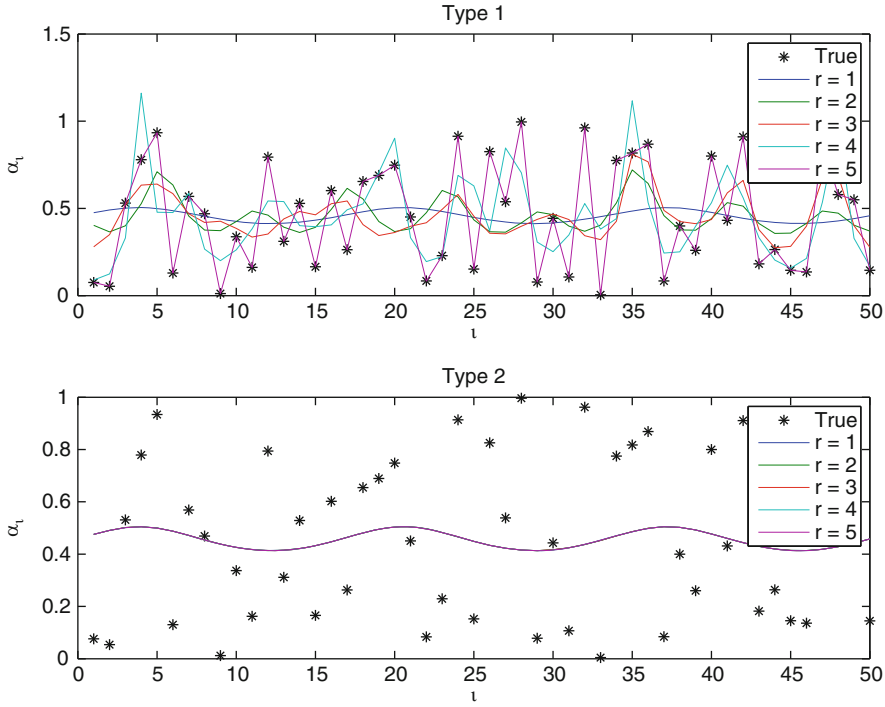


Fig. 5.4 Weight estimation performance comparison using sum of sine waves of Type 1 (*top panel*), or Type 2 (*bottom panel*). The layer has 50 individual EMDs with randomly selected weight values. The curves overlap on *top* of each other for $r = 1,2,3,4,5$ for the *bottom panel*. It is clear that increasing the number of sine waves r improves the estimation using stimulus Type 1, but not Type 2

5.4 Experiments

During experiments, *Eristalis* hoverflies were immobilized with a bee's wax and resin mixture. Recordings were performed intracellularly from Horizontal System (HS) neurons (a type of LPTC) in the left lobula plate using sharp aluminosilicate electrodes pulled on a Sutter P-100 electrode puller (Sutter Instruments, USA). The signal was amplified using a BA-03X amplifier (npi electronic, Germany), with 50 Hz mains disturbance reduced with a HumBug (Quest Scientific, Canada). The signal was digitized at 10 kHz using a NiDAQ 16 bit data acquisition card (National Instruments, NI USB-6210) with the data acquisition toolbox in MATLAB (The MathWorks, USA).

The hoverfly was placed in front of an RGB CRT monitor with a refresh rate of 160 Hz, a mean illuminance of 135 Lx, and a spatial resolution of 640 × 480 pixels (Fig. 5.5a). This corresponds to ca. $100 \times 75\,^{\circ}$ of the hoverfly's field of view. Stimuli were generated with custom software (http://www.flyfly.se) using the psychophysics toolbox (http://psychtoolbox.org) in MATLAB. The stimulus consisted of sinusoidal

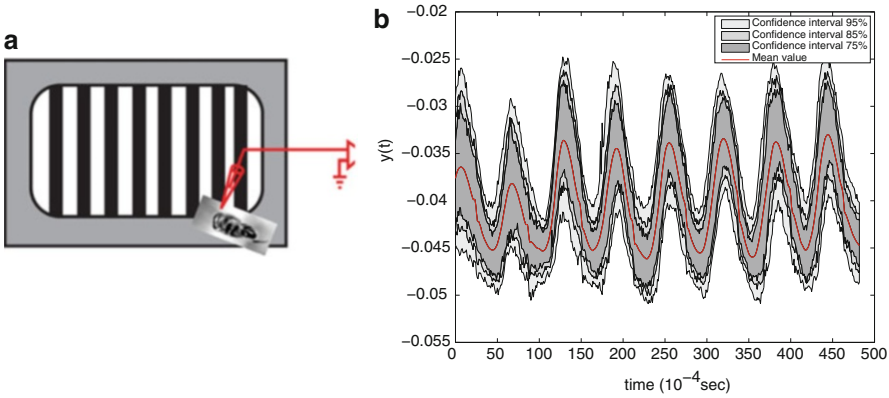


Fig. 5.5 (**a**) Electrophysiological measurement of HS neurons in a hoverfly placed in front of a 160 Hz CRT monitor. Intracellular electrodes were inserted into the left lobula plate while sinusoidal gratings were displayed on the CRT monitor. (**b**) Distribution analysis on the response of the HS neuron over one period of the stimulus (20 Hz). The *bold line* corresponds to the mean value over the time axis, while *grayscale areas* represent the 95, 85, and 75 % confidence regions (Reprinted with permission from [10]. Copyright 2013, AIP Publishing LLC)

full-contrast, full-screen gratings with a wavelength of 50 pixels, moving at 20 Hz. All stimuli were shown for 2 s. Between trials the screen was left at mean luminance for a minimum of 4 s.

5.4.1 Periodicity in the Experimental Data

Figure 5.1 depicts the measured input and output signal of the system. Figure 5.1a shows the time evolution of the luminance at one pixel of the screen measured by a photodiode transducer. The pulsatile high frequency component that is visible throughout the measured data corresponds to the refresh rate of the monitor (160 Hz). Figure 5.1b depicts the measured signal from a HS neuron in response to the visual stimulus in Fig. 5.1a. Apparently, the periodic behavior of the input signal is not faithfully preserved in the measured output. There may be several explanations to this phenomenon. First, the measured neural response represents a pooled output of a population of EMDs with a possibility of other visual input than the experimenter-generated stimuli. Second, other non-visual signals can influence the recorded neural activity.

The measured output is non-periodic but yet dominated by a sequence of pulses that occur approximately at the refresh rate of the screen. In the example here, the same experiment was repeated three times, each run covering 1 s. Thus, in total, 60 realizations of one period of the stimulus were collected. As an illustration of the neural variability, the distribution at each sampling instant are evaluated based on Gaussian distribution, and the result is presented on Fig. 5.5b. The periodicity pattern caused by the CRT monitor's refresh rate is clearly visible in the averaged

LPTC data. The lower amplitude of the first two pulses in the HS mean may indicate the presence of a slow transient component at the beginning of every period of the stimulus.

5.4.2 Identification of a Single EMD

The first stage of the identification algorithm is to estimate the parameters of the linear block in the single EMD model. The linear dynamics are assumed to be of first order. In the identification setup, the photodiode measurement constitutes the input signal to the first channel u^+, while the second input u^- is a delayed version of u^+, see Fig. 5.6a (upper plots). The delay is assumed to be small enough so that u^- vanishes before the next pulse appears at u^+. In this example, the delay is set to 10 sampling instants. An alternative method to find the delay value is gridding over an interval of plausible time delays.

As the output signal, a single pulse from the HS neuron data is considered. As mentioned at the beginning of this chapter, this signal does not correspond to the output of a single EMD but rather to the pooled response of a number of spatially aligned EMDs. For comparison, identification is also performed with the mean data as the output signal. The identification results in Fig. 5.6a (lower plots) show that the estimated model fits the leading edge of the output pulse well but decays faster than the actual data. The estimated transfer functions of the linear blocks are $\hat{W}_R(s) = \frac{3{,}071}{s+426}$ for estimation from raw experimental data and $\hat{W}_M(s) = \frac{2{,}275}{s+392}$ from the mean data.

Figure 5.6b shows how the assumed dynamical order of the linear blocks of the model affects the identification results. The estimated model fits the HS response better with higher order models. However, the pure time-delay model (infinite dimensional) yields the worst fit compared with the first and second order models. Nevertheless, the leading edge of the signal is captured reasonably well.

5.4.3 Identification of a Layer of EMDs

As mentioned earlier, the measured LPTC response corresponds to the pooled output of multiple EMDs. This explains why the estimated single EMD model shows good results at approximating the leading edge of an output pulse but does not capture the rest of the signal. To improve the fit performance, multiple EMDs are considered. The number of contributing EMDs and their weights are evaluated by solving the sparse optimization problem formulated by (5.10).

Figure 5.7 shows the identification results with contribution of multiple EMDs as expressed in model (5.9). The four estimated models are obtained using the first order model and the second order model on raw experimental and mean data. The upper plots depict the fit performance, while the lower plots provide the estimated weights of contributing EMDs. In this example, the shift between the contributing

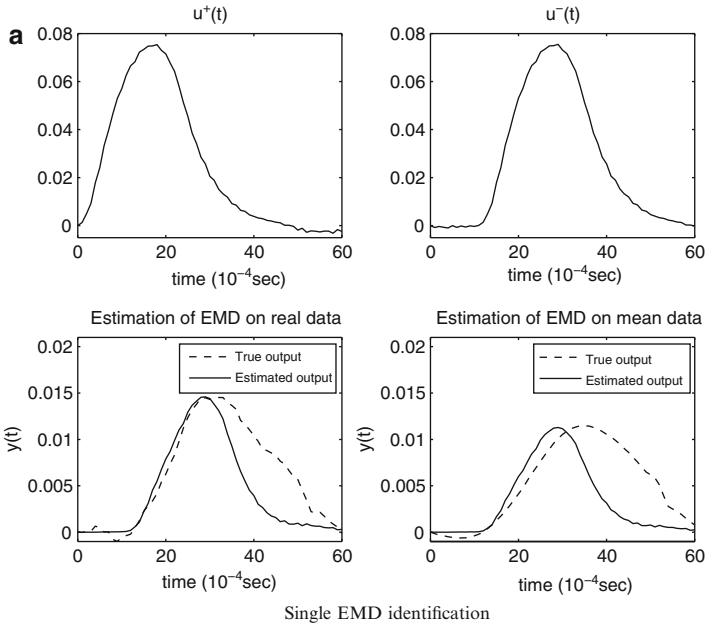
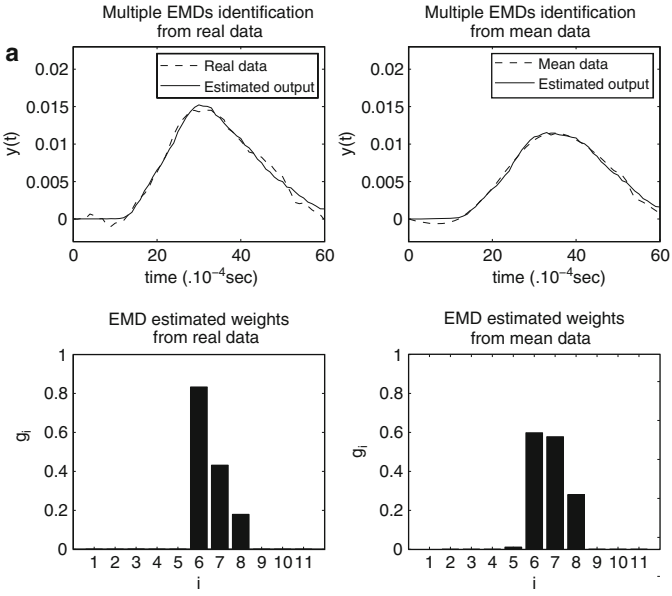
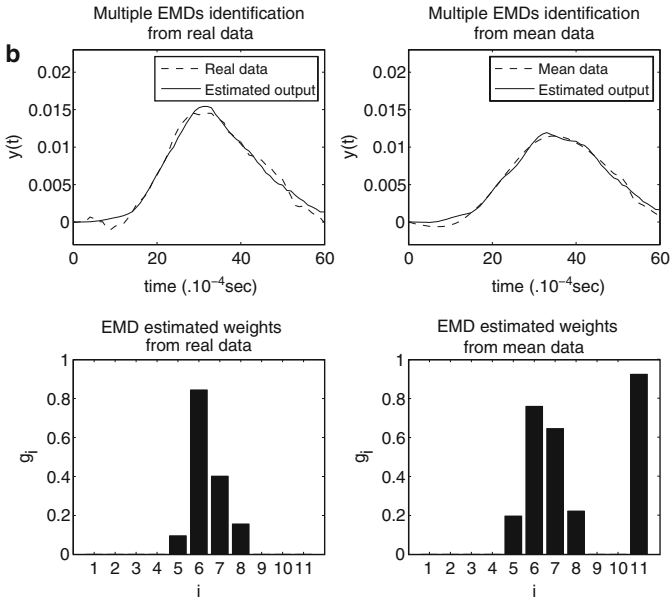


Fig. 5.6 Identification results: (**a**) Identification of the single EMD model. *Upper plot*: The input signals of the EMD model, u^- is a delayed version of u^+. *Lower plots*: identification result for the single EMD model using the measured data and mean data as output signal. (**b**) Comparison of identification results using different models for the linear blocks (Reprinted with permission from [10]. Copyright 2013, AIP Publishing LLC)



Multiple EMDs identification, first order model



Multiple EMDs identification, second order model

Fig. 5.7 Identification results with consideration of multiple EMDs. *Upper plot*: Comparison of the estimated models using the measured data and the mean data. *Lower plot*: The estimated weights of the contributing EMDs evaluated from the measured data and the mean data. (**a**) Using first order model. (**b**) Using second order model (Reprinted with permission from [10]. Copyright 2013, AIP Publishing LLC)

EMDs is assumed to be 10 sampling instants (0.1 ms), which is not biologically feasible, but is suitable for the conceptual model. As it is shown in (5.10), a lower sparsity level of the weights can be achieved at the expense of fit degradation.

For identification with real data, the second order model still gives better fit performance compared to the first order model. However, the difference of the fit performance between first order and second order model for identification on mean data is negligible. This is due to the fact that the mean data obtained through averaging reduce high frequency dynamics. It is also worth mentioning that the positive weights constraint causes poor fit for the first few samples of the data.

Notice that the spatial identifiability analysis of Sect. 5.3.2 applies here only approximately since the actual stimulus to the fly eye is not a sinusoidal grating, but rather a sequence of frame rate pulses whose amplitudes are modulated by the sine wave. Yet, by inspection of the weight plots in Fig. 5.7a, it is easy to conclude that the spatial excitation properties of the modulated sine wave allow for estimation of three spatially distributed EMDs.

Conclusions
The classical EMD model in insect vision has been studied under sinusoidal and pulsatile input. The phenomenon of double-frequency harmonics in the output signal Fourier spectrum in the non-symmetrical EMD model under single tone excitation has been explained. Since the experimentally measured neural response is a result of pooling the outputs of a relatively large number of spatially distributed EMDs, a model for an EMD layer is suggested and estimated by means of a standard sparse estimation algorithm. To enable an unambiguous EMD layer estimate, the spatial excitation properties of stimulating with multiple sinusoidal gratings have been studied. A way of designing such stimuli with a guaranteed spatial excitation order is suggested and validated in a simulation example. A complete EMD layer identification approach based on Laguerre-domain signal representation and sparse optimization has thus been proposed and implemented on laboratory data from hoverfly LPTCs.

Acknowledgements This work was supported by European Research Council via the Advanced Grant 247035 (to EH, AM, SysTEAM), and the Swedish Research Council (to KN, VR 2008-2933).

References

1. A. Borst, M. Egelhaaf, Principles of visual motion detection. Trends Neurosci. **12**, 297–306 (1989)
2. A. Borst, T. Euler, Seeing things in motion: models, circuits, and mechanisms. Neuron **71**, 974–994 (2011)

3. M. Egelhaaf, A. Borst, W. Reichardt, Computational structure of a biological motion-detection system as revealed by local detector analysis in the fly's nervous system. Vis. Res. **20**, 397–407 (1989)
4. B. Fischer, A. Medvedev, L^2 time delay estimation by means of Laguerre functions, in *American Control Conference, ACC'99*, San Diego, June 1999
5. N. Franceschini, A. Riehle, A. Le Nestour, Directionally selective motion detection by insect neurons, in *Facets of Vision*, ed. by D.G. Stavenga, R.C. Hardie (Springer, Berlin/Heidelberg, 1989), pp. 360–390
6. C. Glader, G. Hönäs, P.M. Mäkilä, H.T. Toivonen, Approximation of delay systems – a case study. Int. J. Control **53**, 369–390 (1991)
7. B. Hassenstein, W. Reichardt, Systemtheoretische Analyse der Zeit-, Reihenfolgen- und Vorzeichenauswertung bei der Bewegungsperzeption des Rüsselkäfers Chlorophanus. Zeitschrift für Naturforschung **11**, 513–524 (1956)
8. E. Hidayat, A. Medvedev, Laguerre domain identification of continuous linear time-delay systems from impulse response data. Automatica **48**(11), 2902–2907 (2012)
9. E. Hidayat, A. Medvedev, K. Nordström, Laguerre domain identification of the elementary motion detector model in insect vision, in *11th IFAC International Workshop on Adaptation and Learning in Control and Signal Processing, ALCOSP 2013*, Caen, July 2013
10. E. Hidayat, A. Medvedev, K. Nordström, On identification of elementary motion detectors, in *International Symposium on Computational Models for Life Sciences*, Sydney, Nov 2013, vol. 1559, pp. 14–23
11. A.C. James, D. Osorio, Characterisation of columnar neurons and visual signal processing in the medulla of the locust optic lobe by system identification techniques. J. Compar. Physiol. A **178**(2), 183–199 (1996)
12. P.M. Mäkilä, J.R. Partington, Laguerre and Kautz shift approximations of delay systems. Int. J. Control **72**(10), 932–946 (1999)
13. P.Z. Marmarelis, G.D. McCann, Development and application of white-noise modeling techniques for studies of insect visual nervous system. Kybernetik **12**, 74–89 (1973)
14. J.M. Zanker, M.V. Srinivasan, M. Egelhaaf, Speed tuning in elementary motion detectors of the correlation type. Biol. Cybern. **80**(2), 109–116 (1999)

Chapter 6
Multi-complexity Ensemble Measures for Gait Time Series Analysis: Application to Diagnostics, Monitoring and Biometrics

Valeriy Gavrishchaka, Olga Senyukova, and Kristina Davis

Abstract Previously, we have proposed to use complementary complexity measures discovered by boosting-like ensemble learning for the enhancement of quantitative indicators dealing with necessarily short physiological time series. We have confirmed robustness of such multi-complexity measures for heart rate variability analysis with the emphasis on detection of emerging and intermittent cardiac abnormalities. Recently, we presented preliminary results suggesting that such ensemble-based approach could be also effective in discovering universal meta-indicators for early detection and convenient monitoring of neurological abnormalities using gait time series. Here, we argue and demonstrate that these multi-complexity ensemble measures for gait time series analysis could have significantly wider application scope ranging from diagnostics and early detection of physiological regime change to gait-based biometrics applications.

Keywords Gait time series • Gait-based biometrics • Ensemble classification • Multi-complexity measures • Ensemble decomposition learning • Neurological abnormalities • Minimum spanning tree • Variability analysis • Neurologic disorders • AdaBoost

V. Gavrishchaka (✉)
Department of Physics, West Virginia University, Morgantown, WV 26506, USA
e-mail: gavrishchaka@gmail.com

O. Senyukova
Department of Computational Mathematics and Cybernetics, Lomonosov Moscow State University, Leninskie Gory, 119991 Moscow, Russia
e-mail: osenyukova@graphics.cs.msu.ru

K. Davis
Department of Pathology, University of Michigan, Ann Arbor, MI 48109, USA
e-mail: kristinadavismd@gmail.com

C. Sun et al. (eds.), *Signal and Image Analysis for Biomedical and Life Sciences*,
Advances in Experimental Medicine and Biology 823,
DOI 10.1007/978-3-319-10984-8_6

6.1 Introduction

Development of new technologies has provided inexpensive and unobtrusive means of collecting multi-scale physiological data and led to continuous improvements in clinical instrumentation. Sophisticated portable and wearable systems for real-time collection of physiological data have also become affordable for routine individual use. Increased availability of high-resolution data provides new opportunities for quantitative diagnostics, early abnormality detection, and convenient monitoring.

Analysis techniques compatible with necessarily short time series are essential for many applications. However, it is challenging to construct universal measures or indicators for robust quantification of physiological states from short time series. In express diagnostics, preventive monitoring, and personalization of medical treatment, it is important to find and correctly interpret quantitative measures capable of detecting emerging and transient abnormalities and other subtle regime changes.

Variability analysis of physiological time series provides a generic framework for robust discrimination between normal and abnormal states [14, 20, 32, 35, 39]. The well known application of this methodology is heart rate variability (HRV) analysis approved as one of the modalities for cardiac diagnostics [14, 32, 35, 39]. Compared to traditional electrocardiography (ECG) analysis the method is more robust to noise because it relies only on the interbeat interval signal (RR data) which is very important for analysis of the data from portable and wearable devices. Moreover it is able to detect cardiac and non-cardiac (e.g., emotional) abnormalities lacking well-defined ECG form patterns.

Variability analysis is usually based on nonlinear dynamics (NLD) complexity measures and advanced linear indicators. Unfortunately, the accuracy and stability of such variability measures tend to decrease significantly when the analysis is performed on shorter data segments [11,14,20,32,35,39]. This limitation diminishes the predictive capability of these measures for early detection of both short-lived precursors of emerging physiological regimes and abnormalities with transient patterns.

Recently we have demonstrated that performance of HRV indicators dealing with short time series could be significantly improved through optimal combination of complementary complexity measures, using boosting-like ensemble learning [11, 12]. Such an approach is especially important for early detection of emerging abnormalities and other regime changes where other techniques could often fail.

ECG time series is an important diagnostic and monitoring modality for cardiac abnormalities and related state changes. However, alternative physiological data channels could be informative for other abnormalities. For example, due to advances in technology, gait time series can be easily collected with wearable clinical equipment as well as with general-purpose portable devices such as smartphones where built-in accelerometers are now part of standard configuration. Similar to RR data, measurement of gait stride intervals is very tolerant to noise since only peak-to-peak periods of otherwise complex time series are required.

Variability metrics of gait stride intervals are known to be sensitive to changes in neurological functions associated with aging and development of certain neurological diseases [10, 15, 17, 26]. Long-range correlation and other measures of stride-interval dynamics could be effective in detecting neurological abnormalities and in quantification of their severity [10,15,17,26]. These include Parkinson's (PD) and Huntington's (HD) diseases, amyotrophic lateral sclerosis (ALS), and others.

Remaining challenges in treatment and diagnostics of ALS, PD, HD and other neurological abnormalities maintain significant interest in unobtrusive modalities capable of early diagnostics and robust monitoring of such abnormalities. Therefore, variability indicators computed from stride-interval time series could provide convenient and robust tool for early diagnostics and monitoring of neurological abnormalities. A generic set of NLD complexity measures and linear indicators used in HRV analysis can be directly applied to gait quantification after recalibration.

However, similar to HRV analysis, accuracy of NLD measures and advanced linear indicators could significantly deteriorate when applied to shorter segments of gait time series. Therefore, such indicators would have serious limitations in detecting early intermittent signatures of neurological abnormalities as well as in monitoring effectiveness of medical treatment and its optimization.

In this work we suggest that multi-complexity measures discovered by boosting-like ensemble learning could be effective for early diagnostics and monitoring of neurological abnormalities even when applied to short segments of gait stride intervals. We illustrate validity of our approach using real gait data of normal subjects and patients with ALS, HD and PD. Similarly, we demonstrate capability of multi-complexity ensemble indicators of detecting slow regime changes using gait time series collected from healthy children and teens of different age groups (from 3 to 14 years old). Finally, we present analysis of long gait time series from ten healthy adults suggesting possible application of our meta-indicators in gait-based biometrics. All presented results are based on real gait data available at http://www.physionet.org. Current work provides significant extensions and generalizations of our preliminary results previously reported in [13].

6.2 Variability Analysis of Physiological Time Series: Advantages, Challenges and Multi-complexity Generalization

Majority of NLD measures and linear indicators used for variability analysis require long time series to achieve desired accuracy and stability. For example, many HRV indicators require long time series for stable calculation [14, 32, 39] which could drastically restrict their application scope. Similar restrictions are also relevant for stride-interval analysis. For example, recommendation in one of the recent study [10] is to use segment of at least 600 stride intervals for variability-based diagnostics. Nevertheless, indicators have to be computed on short segments in

order to capture early signs of developing and/or intermittent abnormalities or to detect subtle initial effects of treatment procedures. Indeed, indicator computed on a long time series will average out these short-lived effects and will fail to detect them.

The well-known NLD indicators applicable for HRV and gait analysis are based on detrended fluctuation analysis (DFA) [27], multi-scale entropy (MSEn) [9], and multi-fractal analysis (MFA) including MFA extension of DFA [22]. The discriminative-ability preservation conclusion extends to advanced linear indicators based on power spectrum analysis of the RR time series [35]. We also successfully used power spectrum measure as one of the base indicators for stride-interval analysis.

DFA was proven to be useful in revealing the extent of long-range correlations in time series including HRV applications [27] and diagnostics of neurological abnormalities [10, 15, 17, 26]. First, the investigated time series of length N is integrated. Next, the integrated time series is divided into n boxes. All boxes have the same length. In each box, a least-square line is fitted to the data with y-coordinate denoted by $y_n(k)$ (representing the trend in that box). Finally, the integrated time series, $y(k)$, is detrended as follows:

$$F(n) = \sqrt{\frac{1}{N}\sum_{k=1}^{N}[y(k) - y_n(k)]^2}. \tag{6.1}$$

A linear relationship on the plot of $\log F(n)$ vs. $\log n$ indicates power law (fractal) scaling characterized by a scaling exponent β (slope of the fitted straight line) which is used as HRV indicator.

Multi-scale entropy (MSEn) method [9] has been introduced to resolve limitations of traditional single-scale entropy measures. First, a coarse-graining process is applied to the original time series, x_i. Multiple coarse-grained time series are constructed by averaging the data points within non-overlapping windows of increasing duration, τ:

$$y_j^{(\tau)} = \frac{1}{\tau}\sum_{i=(j-1)\tau+1}^{j\tau} x_i, \tag{6.2}$$

where τ represents the scale factor and $j = 1, \ldots, N/\tau$. The duration of the coarse-grained time series is N/τ. Next, entropy is calculated for each time series and plotted as a function of the scale factor. Different signatures of this function's curve including originally suggested entropy difference between two scales [9] can serve as HRV and other physiological indicators.

HRV indicators based on frequency-domain analysis are often superior in accuracy and stability to the time-domain linear indicators. One of the widely accepted indicators of this type in HRV analysis is a power spectrum ratio of the low-frequency band (0.04–0.15 Hz) to the high-frequency band (0.4–0.15 Hz) [35]. Due to irregularity of the time grid of the RR time series, it is convenient and more

accurate to use a Lomb periodogram for power spectrum calculations instead of Fast Fourier Transform (FFT) [28]. In certain regimes, the accuracy of such power spectrum indicators could be comparable to the best NLD approaches. As discussed in the next section, similar power spectrum measures can also be successfully used as base indicators for stride-interval analysis.

Recently we have illustrated that challenges of variability analysis, when applied to short time series, could be overcome by using a classification framework based on boosting-like ensemble learning techniques that are capable of discovering robust multi-component meta-indicators from a combination of existing variability measures and other incomplete empirical knowledge [11, 12]. Unlike most other combination techniques, the use of boosting is capable of discovering an ensemble of complementary models that has both significantly lower bias (higher accuracy) and lower variance (better stability) compared to each individual model. Potentially more flexible data-driven models (e.g., neural networks) are often unstable due to training data incompleteness, intrinsic non-stationarity, and low signal-to-noise ratio. In addition, such "black-box" systems lack interpretability. In contrast, meta-indicators, discovered by boosting, combine accuracy, stability, and interpretability because they are constructed from the well-understood low-complexity base models.

A typical boosting algorithm such as AdaBoost [4, 30] for the two-class classification problem starts with equal and normalized weights for all training data. Base classifiers, $h_t(x)$, are trained using a weighted error function and the optimal one is chosen at each iteration t. Here x is an input vector. Data points misclassified by the current interation's best model are penalized by the weight-factor adjustment (increase) for the next iteration. Therefore, on each iteration, the algorithm focus is on harder-to-classify samples. The final meta-model, given below, classifies the unknown sample as class $+1$ when $H(x) > 0$ and as -1 otherwise:

$$H_T(x) = \sum_{t=1}^{T} \alpha_t h_t(x). \tag{6.3}$$

Here, the constants α_t are the sequence of combination coefficients obtained, and T is the total number of iterations. Regime adjustments together with important regularization procedures also can be introduced to the original boosting algorithm in several ways [18].

A natural choice of base models could be low-complexity base classifiers, where each of the classifiers uses just one complexity measure, β_i, out of several available choices:

$$y = h(\beta_i[p_i], \gamma). \tag{6.4}$$

Here γ is a threshold level (decision boundary) and p_i is a vector of adjustable parameters of the chosen measure. In our case, β_i may correspond, for example, to either a DFA scaling exponent, a slope of MSEn curve, or a power spectrum ratio. Applying boosting steps to a set of such base classifiers (6.4) with different measures β_i and optimizing over (p_i, γ) on each boosting iteration, we obtain a meta-classifier (6.3).

Additionally, many different physiological regimes, quantified by individual base classifiers from the ensemble, are implicitly encoded in such multi-complexity meta-indicators. In our recent publications [33, 34] we referred to this utilization of ensemble internal structure as ensemble decomposition learning (EDL). We also outlined possible practical application of EDL concept using one of the single-example learning (SEL) frameworks.

The EDL technique can be summarized as follows. Boosting constructs local experts $h_i(x)$ for different implicit regimes or domains of a whole feature space, which ensures good global performance of the final ensemble. Therefore, partial information of wide variety of dynamical regimes becomes implicitly encoded in the obtained ensemble of classifiers. However, only aggregated output is used for normal-abnormal classification, while the rich internal structure of the ensemble is completely ignored. Extraction of this underutilized knowledge could be formalized in terms of ensemble decomposition learning (EDL) [33]. Formally, one can introduce ensemble decomposition feature vector as follows:

$$D(x) = [\alpha_1 h_1(x), \alpha_2 h_2(x), \ldots, \alpha_T h_T(x)]. \tag{6.5}$$

Each sample after ensemble classification procedure can be represented by this vector. Although each individual component of this feature vector may not contain explicit and usable information, collectively, these values may provide detailed and informative state representation of the considered system which is not accessible in the aggregated form given by $H(x)$.

Later, we have also pointed out that, besides particular SEL framework, one can effectively utilize this fine-grain knowledge by using EDL metrics in different types of instance-based learning (IBL) and clustering algorithms including graph-based techniques [34]. For example, we have found that the length change of the minimum spanning tree (MST), constructed using ensemble distance metrics, could provide an early indication of the emerging physiological regimes. All provided illustrations were based on real data for several cardiac abnormalities.

MST representation is motivated by the human perception which organizes information with the most economical encoding. A spanning tree is a connected graph containing all vertices of the original graph without loops, i.e., there exists only one path connecting any two pairs of nodes in the graph. If the edges of the graph are weighted, the spanning tree length is defined as the sum of the weights of its edges. MST is a spanning tree with minimal length among all spanning trees connecting the nodes of the graph. MST of the graph can be derived with Prim's or Kruskal's algorithm [36] with subsequent removal of several longest edges to generate clusters.

Advantages of graph-based representation such as MST have been recently demonstrated in financial applications [23, 25, 37]. Similarly, MST representation can be used to capture essential dependencies and differences between physiological states quantified by the EDL vector. If the length of time series permits computation of N EDL vectors from N consecutive segments of physiological time series, information from $N(N-1)/2$ numbers of distance matrix d_{ij}, the distances between

EDL vectors i and j given by (6.5), defined as l_1 or l_2 norm in T-dimensional space, will be represented with $(N-1)$ edges of MST which can be used for intuitive cluster visualization and analysis. Besides effective clustering, EDL-based MST representation also offers informative aggregated measure such as normalized tree length that could be a sensitive indicator of emerging new regimes or regime changes:

$$L = \frac{1}{N-1} \sum_{d_{ij} \in T} d_{ij}, \tag{6.6}$$

where $(N-1)$ is the number of edges present in MST. Indeed, even in the very early stage when a new regime begins manifesting itself only on short intermittent segments, MST length (6.6) will increase because the distance d_{ij} between EDL vectors of existing and new regimes is significantly higher than the distance between EDL vectors describing the same regime. This could be used for preventive monitoring of healthy subjects as well as for side effects detection in the beginning of new therapy or drug treatment. The described approach is different from EDL-SEL framework mentioned before. Indeed, no EDL vectors associated with particular abnormalities are required and any emerging regimes different from current ones will be detected.

6.3 Diagnostics and Monitoring of Neurological Abnormalities

Several neurologic disorders include abnormalities of movement and gait as prominent symptoms. Huntington's disease is a debilitating and ultimately fatal neurodegenerative disease with an autosomal dominant inheritance pattern. A good family history, symptom history, and physical exam can suggest the diagnosis, which can be then confirmed by molecular tests looking for CAG (codon that codes for the amino acid glutamine) triplet repeats in the Huntington gene [24]. The challenge lies in diagnosing clinical disease onset in an individual who is known to have the genetic abnormality, when classical movement abnormalities may be less prominent [3]. In addition, a key component of active research into potential therapeutic approaches is the ability to monitor symptomatic response in a reliable manner. Currently the United Huntington's Disease Rating Scale is used for this purpose. However, administration of the entire scale is time-consuming. Furthermore, certain items assessed have better discriminatory function than others [38].

Parkinson's disease is another neurologic movement disorder. Diagnosis relies on typical symptoms of motor impairment by physical exam and patient history, after excluding other causes through laboratory and imaging evaluations. A definitive genetic test is not applicable in most cases. Specific medications are available to treat the symptoms of the disease, but making the accurate diagnosis is essential. Patients presenting with gait disturbance as their main initial symptom can have

a significant delay in diagnosis (compared to those who present with a tremor), which in turn delays therapy [8]. Automated gait analysis, as opposed to more subjective clinical assessment, has been shown to be of potential use in diagnosing and following Parkinson's disease [21].

ALS is another progressive neurodegenerative disorder, caused by degeneration of motor neurons which leads to muscle weakness and atrophy. The presenting symptoms depend on which muscles are affected first. Electrophysiological testing, which can be invasive and painful, plays a role in early diagnosis of ALS and differentiation from treatable causes of weakness [5]. However, both in clinical practice and research, a less invasive yet reliable method of assessing disease progression and response to treatment is necessary.

Here we demonstrate that single indicators based on NLD complexity measures could partially preserve their discrimination ability (normal vs. abnormal) even on short segments of gait time series: down to $\approx$100 stride intervals or even shorter. Nevertheless, the reduced accuracy may not be sufficient for many practical applications. However, combination of complementary complexity measures using boosting-like algorithms can significantly increase accuracy and stability of indicators operating on short segments of gait time series. Such multi-complexity measures could be effective for early detection and monitoring of wide range of neurological abnormalities.

To illustrate capabilities of our ensemble-based indicator, we use gait data collected from normal subjects and patients with ALS, HD and PD that are available at http://www.physionet.org. This data set includes gait time series from 15 patients with PD, 20 patients with HD, 13 patients with ALS, and 16 healthy subjects. Each time series consists of up to 300 stride intervals. We use segments as short as 128 stride intervals for calculation of DFA (6.1), MSEn (6.2) and power spectrum measures that are used in base classifiers (6.4). Due to data limitation, total number of segments is increased in the training phase by overlapping. Also, since low-complexity base classifiers are used, we do not find any significant signs of overfitting on out-of-sample data. In the following, performance metrics are computed on all available data.

We should note that previously reported preliminary results [13] were based on indicator calculation from 128 left-only or right-only stride intervals. While such approach is also valid and could be preferable in some cases (e.g., when one leg has certain problems), here we employ the more natural and practical approach by using segments with both left and right stride intervals. This also means that previous results based on 128 left-only or right-only intervals really used time series segments corresponding to 256 stride intervals. In contrast, the current study is based on the shorter segments consisting of 128 intervals.

We apply AdaBoost to discover ensemble meta-indicator given by (6.3) which consists of complementary base classifiers (6.4). The classifier from the first boosting iteration is the best single classifier. In our case, it always happens to be DFA-based classifier. To illustrate discrimination ability of the best single complexity measure, we compute DFA exponents for all 128-interval segments of

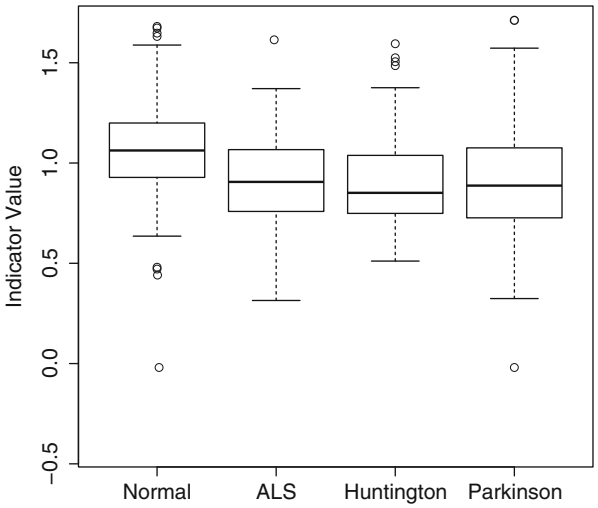


Fig. 6.1 Single DFA measure computed on each of 128-interval segments of stride data from normal control group and patient groups with ALS, HD and PD

stride time series from different groups of patients and summarize them as the box plot in Fig. 6.1. Here we use modified box plot from R with all default settings. Thick horizontal segments represent median, while segments below and above represent 25th and 75th percentiles, respectively. Dashed lines extend to max/min points if they are inside the range obtained by adding the difference between 75th and 25th percentiles scaled by 1.5–75th percentile and subtracting the same value from 25th percentile. Otherwise, points outside this range are plotted as circles.

It is evident from Fig. 6.1 that medians of all abnormal groups (ALS, HD, PD) are clearly below the median of the healthy group. This confirms remaining discrimination ability of a single indicator even for short stride-interval segments. However, there is significant overlapping between normal and abnormal groups even for ranges defined by 25th–75th percentiles. Therefore, normal/abnormal discrimination accuracy of a single measure is quite limited.

However, boosting-based combination of complementary complexity measures could drastically increase accuracy of meta-classifier given by (6.3). Box plot of the aggregated ensemble measure (6.3) computed on 128-interval segments for each group of patients is shown in Fig. 6.2. Now, not only medians of ALS, PD, and HD groups are well below healthy group, but also 25th–75th percentile ranges of all abnormal groups do not overlap with healthy group. Thus, accuracy of the multi-complexity ensemble indicator is significantly increased.

A more formal comparison of the best single measure vs. ensemble indicator is presented in Fig. 6.3. Here we plot detection rates for three reasonable false alarm rates: 10, 20 and 30%. It is clear that boosting-based combination of individual complexity measures can increase detection rate by 40–50%. Such

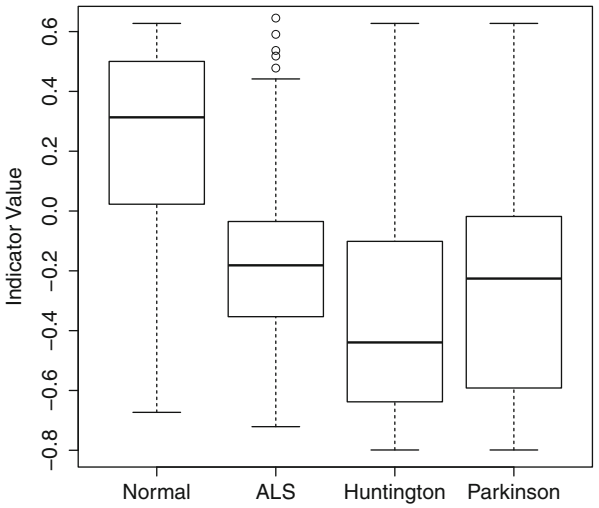


Fig. 6.2 Aggregated ensemble measure computed on each of 128-interval segments of stride data from normal control group and patient groups with ALS, HD, and PD

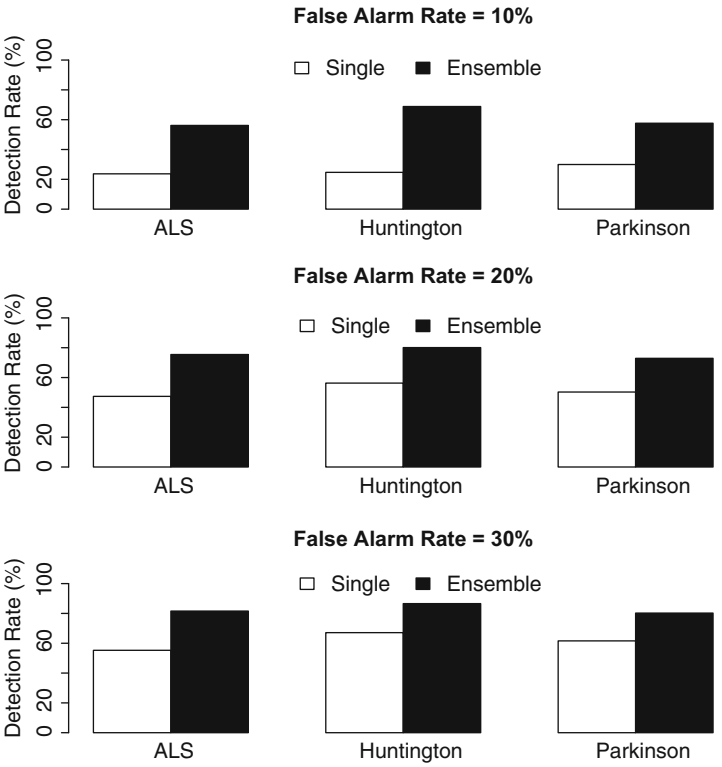


Fig. 6.3 Abnormality detection rates for a given false alarm rate: the best single measure vs. ensemble of multi-complexity measures

dramatic performance improvement suggests that such meta-indicators based on multiple complexity measures could be a reasonable choice for early detection and monitoring of various neurological abnormalities.

6.4 Detection of Emerging Physiological States and Regime Changes

Early detection of emerging physiological state or slow regime changes is often more challenging than accurate diagnostics of the developed abnormality. Indeed, many discriminative features used in diagnostics are not yet present when abnormality is in early development stage. Similar challenges are also typical for early detection of changes during personalization of drug treatment or therapy.

High detection rates with acceptable false alarm rates shown in Fig. 6.3 illustrate ability of the multi-complexity meta-classifier to discriminate between gait time series from normal subjects and subjects with various developed abnormalities. Implicitly, this also suggests potential ability to detect early signs of the developing abnormality and other regime changes. However, direct illustration based on data with emerging or intermittent pathologies could be more convincing. While we are not aware of any large open-access databases capturing slow development of neurological abnormalities, other gait databases can be used for illustration of slow physiological regime changes. One of them is gait maturation database first analyzed by Hausdorff et al. [16] and now available at http://www.physionet.org.

Gait maturation database is a collection of gait time series from 50 children of various age groups: from 3 to 14 years old. For each subject, time series is up to 500 stride-intervals long. It is known that in very young children, immature control of posture and gait results in unsteady locomotion [16]. In children $\approx$3 years old, gait appears relatively mature. However, as suggested in [16], the dynamics of walking changes continues beyond this age. This was confirmed by quantitative analysis of 50 children from gait maturity database [16]. Single time- and frequency-domain measures as well as DFA-based measures have been used in that study. It was demonstrated that, while gait in younger age groups resembles that of adults with neurological abnormality, it continuously matures and approaches the dynamical range of healthy young adults as age increases.

In the analysis of gait maturation data, Hausdorff et al. [16] calculated indicators using significantly long segments (at least 256 stride intervals), and there was still wide overlap of indicator values among different age groups. Such overlap could only increase for shorter segments. This overlap is not critical for the main objective of the analysis presented in [16]. Indeed, the authors demonstrated statistically significant trend of gait characteristics approaching those of healthy adults when children age increases. However, for early detection of slow regime change due to developing abnormality or initial treatment effects, insufficient discrimination of single indicators could make them useless in practice.

Thus, gait maturation database offers convenient real-life data to demonstrate advantages of multi-complexity ensemble measures over single indicators in detect-

ing emerging physiological states and slow regime changes. For this purpose, we compare the best single indicator (DFA) and ensemble of multi-complexity measures discovered in the normal-abnormal classification problem in the previous section. It is important to note that none of the gait maturation data were used in the training phase.

We applied these indicators to short (128-interval) segments from different age groups and summarized results as box plots in Figs. 6.4 and 6.5.

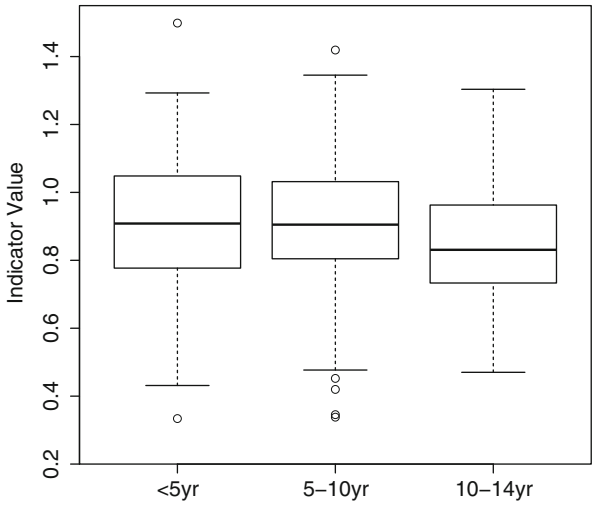


Fig. 6.4 Single DFA measure computed on each of 128-interval segments of stride data from three different age groups of healthy subjects

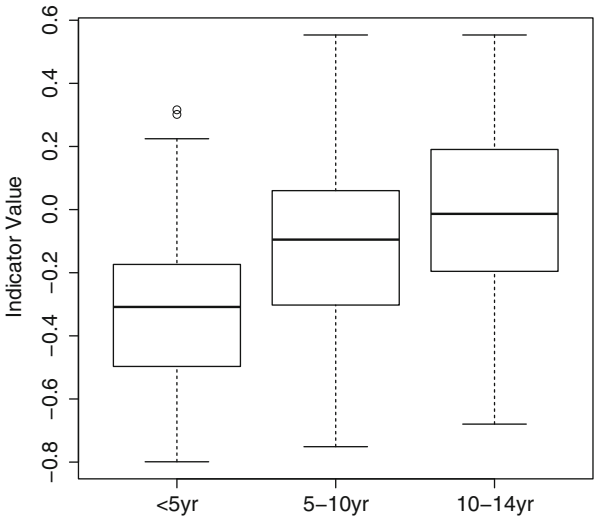


Fig. 6.5 Aggregated ensemble measure computed on each of 128-interval segments of stride data from three different age groups of healthy subjects

Figure 6.4 demonstrates that single DFA indicator is not capable to detect any clear trend in gait dynamics evolution as children age increases. On the other hand, as evident from Fig. 6.5, aggregated output, $H(x)$, of the multi-complexity ensemble indicator shows clear trend towards gait dynamics of healthy adults as age increases. Remember that the range of probability-like output, $H(x)$, is $[-1, +1]$, with healthy state corresponding to positive numbers. Therefore, aggregated output of multi-complexity ensemble measures could be a promising metric for early detection of subtle and/or slow changes in gait dynamics.

6.5 Biometrics Based on Gait Time Series Analysis

In the previous section, we have demonstrated that multi-complexity ensemble measures, obtained as binary classifiers for normal-abnormal diagnostics problem, could have much wider application scope. Indeed, the aggregated ensemble output could be a sensitive and robust indicator of physiological regime changes. In addition, as shown previously in the context of cardiac abnormalities, multi-complexity ensembles could be also used for multi-class discrimination and rare-state quantification using EDL representation [33, 34]. Similar approach could be also effective for physiological state discrimination based on gait time series. Gait-based biometrics is an example of such practical problem where effective discrimination among multiple physiological states or classes is crucial.

Recently, gait recognition has become a topic of interest within the computer vision applications, due to its growing importance as a biometric modality [2, 6, 29]. An important motivation for gait recognition research has been provided by psychophysical experiments with Moving Light Displays (MLDs) pioneered by Johansson [19]. Johansson's experiments demonstrated the ability of humans to recognize the type of movement of a person solely from observing the 2D motion pattern generated by light bulbs attached to the person. Similar experiments indicated that even the identity of a familiar person, as well as the gender of the person, might be recognizable from MLDs [1]. These experiments provided insight into motion perception in the human visual system and suggested that motion patterns generated by the human gait encode information that could be unique for the moving person.

The fact that each person seems to have a distinctive (idiosyncratic) way of walking is hardly surprising from a biomechanics standpoint [40]. Human ambulation consists of synchronized integrated movements of hundreds of muscles and joints in the body. Although these movements follow the same basic pattern for all humans, they seem to vary from one individual to another in certain details such as their relative timing and magnitudes. Much research in biomechanics and clinical gait analysis is devoted to the study of the inter-person and intra-person variability of gait, mainly to determine normal vs. pathological ranges of variation. There is an increased interest in gait as a biometric, mainly due to its non-intrusive and non-concealable nature as well as possibility of remote biometrics. Considerable research efforts are being devoted in the computer vision community to characterize

and extract gait dynamics automatically from video [2, 6, 29]. Biometric systems for human identification at a distance are in increasing demand in various real-life applications. Many biometric modalities such as face recognition, iris, fingerprints, palm prints, and hand geometry have been systematically studied and employed in many practical systems. However, these approaches suffer from two main disadvantages: (1) high failure rate when only low resolution images and pictures taken at a distance are available, and (2) necessity of subject cooperation for accurate results. For these reasons, innovative biometric recognition methods for human identification at a distance have significant potential and appeal for surveillance and forensic applications [6].

The most common categories of gait recognition, currently used and discussed in the literature, are appearance-based and model-based approaches [2, 29]. Among the two, the appearance-based approaches suffer from changes in the appearance owing to the change of the viewing or walking directions [2, 29]. Model-based approaches extract the motion of the human body by means of fitting their models to the input images. Model-based approaches are view- and scale-invariant as well as take into account the kinematic characteristics of walking manner [2,29]. In general, a gait is considered as being composed of a sequence of kinematic characteristics of human motion and most systems in existence recognize it by the similarity of these characteristics.

However, for accurate recognition, both approaches require significant amount of details to be extracted from sequential video images. Unfortunately, in many practical cases the quality of available video may be poor due to insufficient lighting, covering of motion details by cloth, and other reasons. In such circumstances, one of the gait characteristics that could still be extracted is time period between consecutive steps. This is similar to RR intervals that could be quite accurately measured even from very noisy ECG time series and still used for cardiac diagnostics unlike ECG waveforms. Therefore, multi-complexity ensemble measures for gait time series analysis could be potentially employed as complementary approach in biometric applications, especially in cases where quality of the video prohibits usage of more traditional techniques.

Besides remote gait-based biometrics from video sequences, gait recognition from accelerometer data, that are currently available in all standard smartphones and other wearable devices, becomes comparably important [7]. Indeed, remote acquisition of accelerometer time series could be used to identify the person currently carrying the phone and in related applications. Until now, the most common approaches to analyze such time series included direct feature extraction using wavelet transform or similar techniques with subsequent application of standard classification algorithms [7]. However, when quality of time series is poor, it would be beneficial to extract the most robust feature such as stride interval time series with subsequent application of suitable techniques for further analysis. Again, multi-complexity measures discussed in this paper could be applicable to this set of gait-based biometric problems.

Success of multi-complexity measures in normal-abnormal classification does not warranty its effectiveness in gait-based biometrics. Indeed, stride-interval

dynamics is quite different in healthy subjects and patients with neurological abnormalities. This explains success in binary "normal-abnormal" classification. However, differences between same-class (e.g., healthy) subjects could be much more subtle. Also, here we are dealing with multi-class classification problem that is often significantly more challenging than binary (2-class) classification. Nevertheless, our preliminary results presented below indicate that multi-complexity ensemble measures could differentiate between different individuals using their stride interval time series. This suggests applicability of our approach to gait-based biometrics problems discussed above. Our results are based on the analysis of gait time series from ten healthy subjects available at http://www.physionet.org. This database consists of 3.5×10^4 stride intervals.

Multi-class classification algorithms are generally based on the reductions to the binary case. A common choice, applicable to both single classifiers and boosting algorithms, is "one-against-all" reduction where separate classifiers for each of the considered classes are built [31]. After test sample is presented to such set of classifiers, the sample is considered belonging to the class for which probability-like output of the corresponding classifier is the largest among all available classifiers.

First, we demonstrate possibility of discrimination among subjects based on ensembles of multi-complexity measures that are built using "one-against-all" approach. We apply our framework to build ten ensemble classifiers given by (6.3) corresponding to each of the ten healthy subjects. In the test phase, we present data from each of the subjects to this set of ten classifiers. If multi-complexity measures are capable to discriminate among different subjects, the maximum output (up to $+1$) should be obtained from the classifier corresponding to the subject whose data are currently presented. In other words, the output differences between the classifier corresponding to the subject of the presented data and all other classifiers (total, 9 numbers) should be positive with values up to $+2$. This is exactly what we observe in Fig. 6.6 where results of our multi-class classification test are presented. Indeed, all data points, except one, are positive numbers. Therefore, these preliminary results suggest effectiveness of multi-complexity measures in the context of "one-against-all" multi-class classification that can be used in gait-based biometrics.

While "one-against-all" multi-class classification is valid approach to gait-based biometrics, it has important limitations. Indeed, it is applicable only to limited number of classes (individuals) since the data from each class (individual) should be available in training phase to build classifiers for each class. Therefore, other individuals that are not present in training data cannot be identified. However, in many important practical cases one would be interested to measure the similarity between two gait time series to conclude whether these time series belong to the same individual or not. Such similarity metrics should be robust enough to work for individuals that are not included in the training phase. This is possible if, instead of aggregated outputs of boosting-based classifiers, one uses EDL vectors (6.5) of the ensemble classifiers.

Here we describe and test one of the possible EDL-based frameworks for gait-based biometrics. Previously, in the context of cardiac abnormalities, we have shown that normalized MST tree length (6.6) computed from distance matrix of

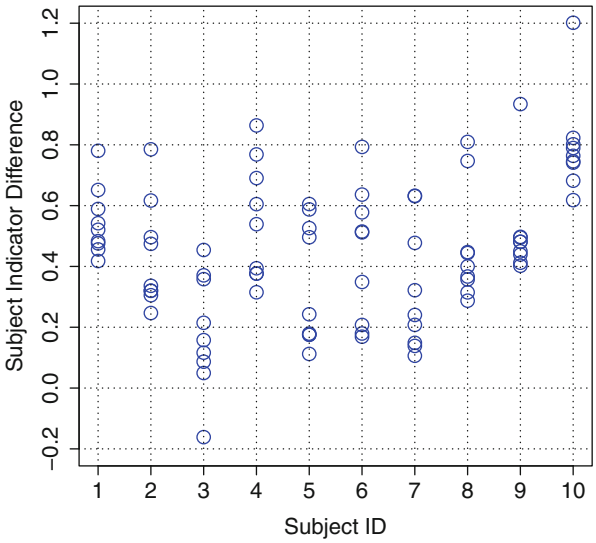


Fig. 6.6 Difference between median of the aggregated outputs of the current subject classifier and medians of outputs from other subjects classifiers computed on 128 stride interval segments. All classifiers are computed using current subject data. Calculations repeated for each ten healthy subjects

consecutive EDL vectors could be used as sensitive indicator of emerging patterns and slow regime changes [34]. The same approach can be used in EDL-based biometrics with slight modification. Here, instead of using distance metrics of EDL vectors from the same individual, we can create cross-subject distance (proximity) matrix where distances between all EDL vectors of one subject (individual) and all EDL vectors of another subject (individual) are computed. Then we still apply MST technique for low-dimensional representation of such matrix and compute MST tree length as an aggregated measure of EDL-based distance or proximity measure between two individuals.

Operationally, such EDL-based biometrics framework can be implemented as follows. Given gait time series from N subjects (individuals), we build N multi-complexity ensembles as in "one-against-all" multi-class classification approach. Then, we combine N EDL vectors of these classifiers to obtain one multi-feature EDL vector. Such EDL vector can be applied not only to N subjects used in training but to quantify proximity of any two gait time series. Indeed, even for relatively small number of individuals used in training, their EDL vectors provide rich multi-feature (multi-regime) representations that can be generically applied to quantify any gait time series.

For demonstration of the effectiveness of such EDL-based approach, we apply it to the same gait database of ten healthy individuals as in "one-against-all" multi-class classification discussed earlier. We combine ten EDL vectors into a single EDL vector that should capture large number of different micro-regimes and their

combinations. In testing phase, we compute consecutive EDL vectors from gait time series of a subject that needs to be identified. Then, we compare these consecutive EDL vectors with EDL vectors of reference subjects by construction MST tree from cross-subject proximity matrix and computing MST tree length (6.6). The lowest MST tree length among reference subjects will identify the tested subject. Please note that the set of reference subjects can be increased without retraining for obtaining new EDL vector.

In our series of tests we use gait data from each of ten subjects for EDL-based identification. In all ten tests, the minimum MST tree length correctly identifies tested individual. Details of six out ten tests are presented in Fig. 6.7. We see that

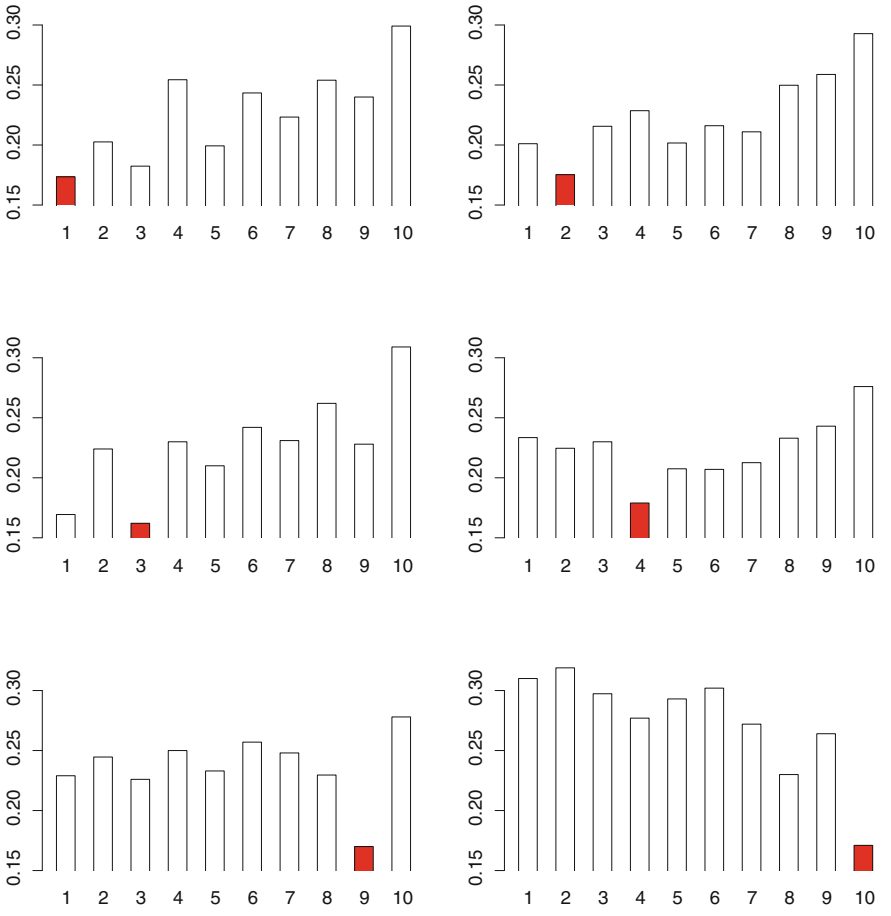


Fig. 6.7 Normalized MST length computed from the cross-subject proximity matrix. Proximity matrix is defined as l_2 distance between all pairs of EDL feature vectors computed on 128 stride interval segments. *Red column* indicates current reference subject whose proximity to other subjects including himself is measured by MST length

the lowest MST tree length points to the correct reference subject, i.e., if data from the n-th subject is presented, the lowest MST tree length will be obtained from cross-subject proximity matrix with the n-th reference individual. Thus, our results provide preliminary indication that generic high-dimensional feature vector can be constructed using training data from reasonable number of individuals and used in the wide range gait-recognition applications without any additional retraining.

Conclusions

We have demonstrated that boosting-based combination of multi-complexity measures could significantly improve quantitative analysis of short gait time series and could be applied for early detection of neurological abnormalities and their monitoring. Our conclusions are illustrated on real gait data from healthy subjects and patients with several neurological abnormalities. We have also demonstrated that multi-complexity ensemble measures can be effective in early detection of slow physiological regime changes and in gait-based biometric applications. More detailed analysis of our ensemble-based multi-complexity indicators on larger gait data sets is warranted.

References

1. C. Barclay, J. Cutting, L. Kozlowski, Temporal and spatial factors in gait perception that influence gender recognition. Percept. Psychophys. **23**(2), 145–152 (1978)
2. C. BenAbdelkader, R.G. Cutler, L.S. Davis, Gait recognition using image self-similarity. EURASIP J. Appl. Signal Process. **4**, 572–585 (2004)
3. K.M. Biglan, Y. Zhang, J.D. Long, M. Geschwind, G.A. Kang, A. Killoran, W. Lu, E. McCusker, J.A. Mills, L.A. Raymond, C. Testa, J. Wojcieszek, J.S. Paulsen, Refining the diagnosis of Huntington disease: the PREDICT-HD study. Front. Aging Neurosci. **5**, 12 (2013)
4. C.M. Bishop, *Pattern Recognition and Machine Learning* (Springer, Heidelberg, 2006)
5. P. Bouche, N. Le Forestier, T. Maisonobe, E. Fournier, J.C. Willer, Electrophysiological diagnosis of motor neuron disease and pure motor neuropathy. J. Neurol. **246**(7), 520–525 (1999)
6. I. Bouchrika, M. Goffredo, J. Carter, M. Nixon, On using gait in forensic biometrics. J Forensic Sci. **56**(4), 882–889 (2011)
7. M. Boyle, A. Klausner, D. Starobinski, A. Trachtenberg, H. Wu, Gait-based smartphone user identification, in *Proceedings of the 9th International Conference on Mobile systems, Applications, and Services (MOBISYS'11)*, Bethesda (2011), pp. 395–396
8. D.P. Breen, J.R. Evans, K. Farrell, C. Brayne, R.A. Barker, Determinants of delayed diagnosis in parkinson's disease. J. Neurol. **260**(8), 1978–1981 (2013)
9. M. Costa, A.L. Goldberger, C.-K. Peng, Multiscale entropy analysis of biological signals. Phys. Rev. E **71**, 021906 (2005)
10. S. Damouras, M.D. Chang, E. Sejdic, T. Chau, An empirical examination of detrended fluctuation analysis for gait data. Gait Posture **31**, 336–340 (2010)
11. V.V. Gavrishchaka, O. Senyukova, Robust algorithmic detection of the developed cardiac pathologies and emerging or transient abnormalities from short periods of RR data, in *International Symposium on Computational Models for Life Sciences, Oct 2011*, Toyama City. AIP Conference Proceedings, vol. 1371 (AIP Publishing, New York, 2011), pp. 215–224

12. V.V. Gavrishchaka, O. Senyukova, Robust algorithmic detection of cardiac pathologies from short periods of RR data, in *Knowledge-Based Systems in Biomedicine and Computational Life Science*, ed. by T.D. Pham, L.C. Jaim. Studies in Computational Intelligence, vol. 450 (Springer, Heidelberg, 2013), pp. 137–153
13. V. Gavrishchaka, K. Davis, O. Senyukova, Multi-complexity measures for early detection and monitoring of neurological abnormalities from gait time series, in *2013 International Symposium on Computational Models for Life Sciences (CMLS-13), Sydney, Nov 2013*. AIP Conference Proceedings, vol. 1559 (AIP Publishing, New York, 2013), pp. 47–56
14. L. Glass, J. Milton, *Dynamical Disease: Mathematical Analysis of Human Illness* (AIP, New York, 1995)
15. J.M. Hausdorff, S.L. Mitchell, R. Firtion, C.K. Peng, M.E. Cudkowicz, J.Y. Wei, A.L. Goldberger, Altered fractal dynamics of gait: reduced stride interval correlations with aging and Huntington's disease. J. Appl. Physiol. **82**(1), 262–269 (1997)
16. J.M. Hausdorff, L. Zemany, C. Peng, A.L. Goldberger, Maturation of gait dynamics: stride-to-stride variability and its temporal organization in children. J Appl. Physiol. **86**(3), 1040–1047 (1999)
17. J.M. Hausdorff, A. Lertratanakul, M.E. Cudkowicz, A.L. Peterson, D. Kaliton, A.L. Goldberger, Dynamic markers of altered gait rhythm in amyotrophic lateral sclerosis. J. Appl. Physiol. **88**, 2045–2053 (2000)
18. R. Jin, Y. Liu, L. Si, J. Carbonell, A.G. Hauptmann, A new boosting algorithm using input-dependent regularizer, in *20th International Conference on Machine Learning*, Washington DC (AAAI, Palo Alto, 2003)
19. G. Johansson, Visual perception of biological motion and a model for its analysis. Percept. Psychophys. **14**(2), 201–211 (1973)
20. H. Kantz, T. Schreiber, *Nonlinear Time Series Analysis* (Cambridge University Press, Cambridge, 1997)
21. J. Klucken, J. Barth, P. Kugler, J. Schlachetzki, T. Henze et al., Unbiased and mobile gait analysis detects Motor Impairment in Parkinson's disease. PLoS ONE **8**(2), e56956 (2013). doi:10.1371/journal.pone.0056956
22. D. Makowiec, A. Dudkowska, M. Zwierz, R. Galaska, A. Rynkiewicz, Scale invariant properties in heart rate signals. Acta Physica Polonica **B37**(5), 1627–1639 (2006)
23. R.N. Mantegna, Hierarchical structure in financial markets. Eur. Phys. J. B **11**, 193–197 (1999)
24. M.J.U. Novak, S.J. Tabrizi, Huntington's disease: clinical presentation and treatment, in *International Review of Neurobiology*, ed. by J. Brotchie, E. Bezard, P. Jenner (Academic, Chicago, 2011), pp. 297–323
25. J.-P. Onnela, A. Chakraborti, K. Kaski, J. Kertesz, A. Kanto, Dynamics of market correlations: taxonomy and portfolio analysis. Phys. Rev. E **68**, 056110 (2003)
26. L. Ota, H. Uchitomi, K. Suzuki, M.J. Hove, S. Orimo, Y. Miyake, Relationship between fractal property of gait cycle and severity of Parkinson's disease, in *IEEE/SICE International Symposium on System Integration (SII)*, Kyoto, Dec 2011, pp. 236–239
27. C.-K. Peng, S. Havlin, E.H. Stanley, A.L. Goldberger, Quantification of scaling exponents and crossover phenomena in nonstationary heartbeat time series. Chaos **5**, 82–87 (1995)
28. W.H. Press, S.A. Teukolsky, W.T. Vetterling, B.P. Flannery, *Numerical Recipes in C: The Art of Scientific Computing* (Cambridge University Press, Cambridge, 1992)
29. M.P. Rani, G. Arumugam, An efficient gait recognition system for human identification using modified ICA. Int. J. Comput. Sci. Inf. Technol. (IJCSIT) **2**(1), 55–67 (2010)
30. R.E. Schapire, *The Design and Analysis of Efficient Learning Algorithms* (MIT, Cambridge, 1992)
31. R.E. Schapire, Y. Freund, *Boosting: Foundations and Algorithms* (MIT, Cambridge, 1992)
32. A.J.-E. Seely, P.T. Macklem, Complex systems and the technology of variability analysis. Crit. Care **8**, 367–384 (2004)
33. O. Senyukova, V. Gavrishchaka, Ensemble decomposition learning for optimal utilization of implicitly encoded knowledge in biomedical applications, in *IASTED International Conference on Computational Intelligence and Bioinformatics*, Pittsburgh (ACTA, Calgary, 2011), pp. 69–73

34. O. Senyukova, V. Gavrishchaka, M. Koepke, Universal multi-complexity measures for physiological state quantification in intelligent diagnostics and monitoring systems, in *Proceedings of The First International Aizu Conference on Biomedical Informatics and Technology (ACBIT 2013), Sept 2013*, Aizu-Wakamatsu. Communications in Computer and Information Science, vol. 404 (Springer, Heidelberg, 2013), pp. 69–83
35. Task Force of the European Society of Cardiology and the North American Society of Pacing and Electrophysiology, Heart rate variability: standards of measurement, physiological interpretation, and clinical use. Circulation **93**, 1043–1065 (1996)
36. S. Theodoridis, K. Koutroumbas, *Pattern Recognition* (Academic, Boston, 1999)
37. M. Tumminello, F. Lillo, R.N. Mantegna, Correlation, hierarchies, and networks in financial markets. J. Econ. Behav. Organ. **75**, 40–58 (2010)
38. A.L. Vaccarino, K. Anderson, B. Borowsky, K. Duff, J. Giuliano, M. Guttman, A.K. Ho, M. Orth, J.S. Paulsen, T. Sills, D.P. van Kammen, K.R. Evans, PREDICT-HD and REGISTRY investigators coordinators: an item response analysis of the motor and behavioral subscales of the unified huntington's disease rating scale in huntington disease gene expansion carriers. Mov. Disord. **26**(5), 877–884 (2011)
39. A. Voss, S. Schulz, R. Schroederet, M. Baumert, P. Caminal, Methods derived from nonlinear dynamics for analysing heart rate variability. Philos. Trans. R. Soc. A **367**, 277–296 (2008)
40. D. Winter, *The Biomechanics and Motor Control of Human Gait* (Univesity of Waterloo Press, Waterloo, 1987)

Chapter 7
Development of a Motion Capturing and Load Analyzing System for Caregivers Aiding a Patient to Sit Up in Bed

Akemi Nomura, Yasuko Ando, Tomohiro Yano, Yosuke Takami, Shoichiro Ito, Takako Sato, Akinobu Nemoto, and Hiroshi Arisawa

Abstract This research was carried out to analyze the actions of caregivers when aiding a patient to sit up in bed. The new system showed that three dimensional analysis could be performed even on points on the subjects' bodies that were hidden from view. We also developed a method to estimate the load on the lumbar region of caregivers based on the kinetic analysis of the human body. Using this system we were able to evaluate the performance of both lay and professional caregivers. We found a clear difference between the performances of the two types of caregivers, and noted that the professional adopted a posture that was safe and did not stress the lumbar vertebrae, whereas the layperson tended to adopt an unsafe posture.

Keywords Motion capture system • Kinetic analysis • Kinematic analysis • Link-segment human model • Care for patients to sit up

7.1 Introduction

Because of the ever-growing number of aged people needing nursing care in Japan and elsewhere, there is a pressing need to reduce the burdens imposed upon their caregivers. To prevent such patients from being confined to their bed, it is necessary to make it possible for them to use wheelchairs. From the point of view of information engineering, the human body is a multi-jointed entity with a high degree

A. Nomura (✉) • Y. Ando • A. Nemoto
Yokohama City University, 3-9, Fukuura, Kanazawa-ku, Yokohama, Japan
e-mail: gnomu@yokohama-cu.ac.jp; ando@yokohama-cu.ac.jp; anemoto@med.yokohama-cu.ac.jp

T. Yano • Y. Takami • S. Ito • T. Sato • H. Arisawa
Yokohama National University, 79-7, Tokiwadai, Hodogaya-ku, Yokohama, Japan
e-mail: yano@arislab.ynu.ac.jp; takami@arislab.ynu.ac.jp; ito@arislab.ynu.ac.jp; sugar@arislab.ynu.ac.jp; arisawa@ynu.ac.jp

C. Sun et al. (eds.), *Signal and Image Analysis for Biomedical and Life Sciences*, Advances in Experimental Medicine and Biology 823,
DOI 10.1007/978-3-319-10984-8_7

of freedom when taking into account the patient and the caregiver. It is very difficult to design a system based on kinematics and kinetics to reduce the caregiver's burden.

The importance of evaluating the acts of caregiving is well recognized, and the load imposed on the lumbar region when moving a patient from bed to wheelchair has been analyzed biometrically using motion capture techniques [2, 3, 5, 7]. However, since the patient and caregiver are in intimate contact during this act, previous studies have not been able to analyze the act in three dimensions. Such studies have used only two-dimensional models, without taking into account the rotation and lateral bending of the bodies [1, 4, 8–10]. These factors have not been studied with a three-dimensional model. To do a three-dimensional analysis, a motion capture system employing several cameras is needed, and a number of markers must be applied to the subjects. Furthermore, analysis cannot be done if the markers become hidden.

For this reason, our aim has been to observe helping a patient sit up in bed and to develop a system that permits three-dimensional analysis of even parts of the scene that are not visible, and further, to use the system to evaluate the load imposed upon a nurse and an adult layperson when performing this action. In this chapter we describe a combination of motion capture apparatus with multifunction sensors (combination of tri-axial acceleration sensor and geomagnetic sensor) which we used to aid a patient to sit up in bed and move to a wheelchair.

7.2 The Fundamental Principals of Using Motion Capture to Analyze the Parts of an Action not Visible to the Camera

The actions of standing and sitting have been analyzed by existing optical motion capture methods. To do this a number of markers must be attached to the body of the subject, and at least two cameras must be set at different angles and synchronized to take a rapid series of frames. The position of each marker is then obtained by a Direct Linear Transformation method and analysis is performed. This is effective in actions like standing up, where most of the markers are not hidden. In actions like moving a patient or aiding a patient to sit up, the bodies of the caregiver and the patient are very close to each other. In such cases many of the markers are hidden and the motion capture system cannot work effectively.

In order to analyze the movements of two human bodies even if they are close, we propose the following system. Figure 7.1 illustrates the concept of the system. With this system, the actions of the caregiver are analyzed. The caregiver puts his/her upper arm deeply under the back of a patient lying face upward and aids the patient to sit up. At this time the caregiver's arms are not visible to the camera, and a load is imposed upon the arms.

In order to analyze this hidden area, we devised the following method. First we construct a link-segment model consisting of 16 segments and 14 joints. The length, width, and points of rotation of each part are identified on the body of the caregiver.

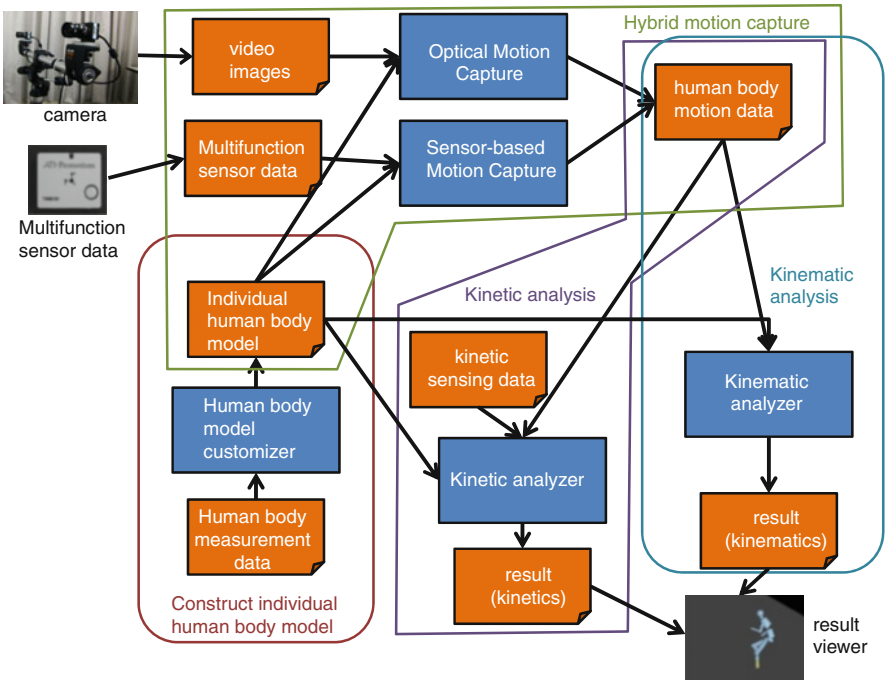


Fig. 7.1 Development of a physical motion and load analysis system

Using the link-segment model we attach markers to all parts except the upper arm and the forearm. On both right and left sides we attach multifunction sensors to the parts of the upper arm and forearm where the sensors will not interfere with the patient during performance of the action. We will give the specifications of the multifunction sensors later. This condition is shown in Fig. 7.2.

We determine the three-dimensional location (on the model) of the markers from the trunk to the shoulders, on the legs, and on the head. Regarding the upper arms and forearms, we estimate the spatial location of the markers (three degrees of freedom) using data from the multifunction sensors (Fig. 7.3). This is possible because the size of the body joints and their location on the body are known. The concept of our measuring system is shown in Fig. 7.4.

In this way the markers attached to the parts of the body model, exclusive of the upper arm and forearm on both right and left sides, can be determined (the markers are recorded with a motion-sensitive camera, Fig. 7.5). Then, if the upper arm segment that rotates around the shoulder joint and the forearm segment that rotates around the elbow joint are estimated and their location in space is determined by their rotation around the three axes of X, Y, and Z (i.e., three degrees of freedom), a three-dimensional (six degrees of freedom) position can be determined. We attempted to do this with multifunction sensors that are able to sense the direction of the magnetic field of the earth.

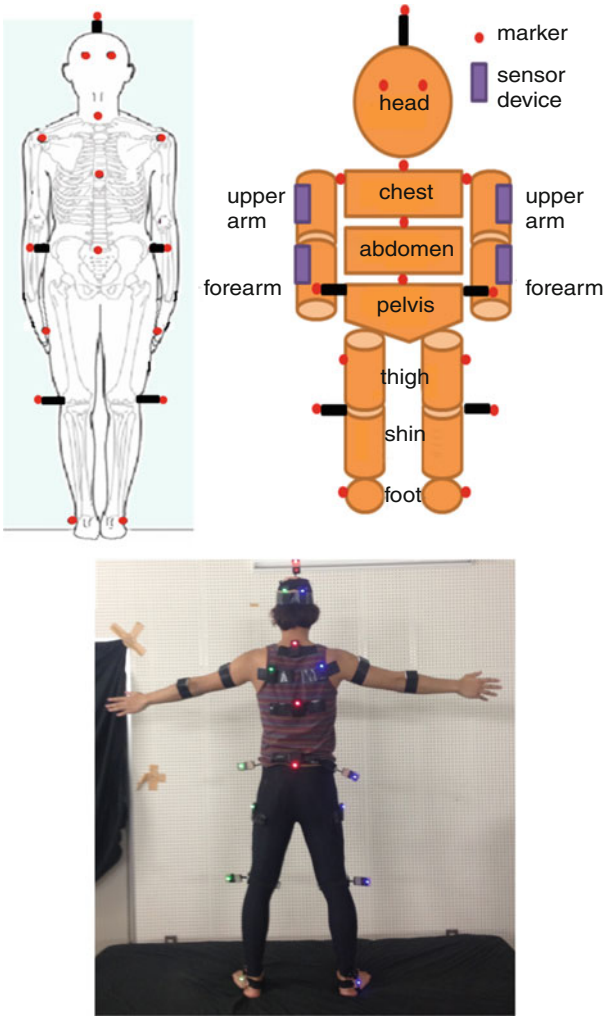


Fig. 7.2 The multifunction sensors to the upper body

Figures 7.6 and 7.7 illustrate this principle. Because gravitational acceleration is always in a vertical direction, a multifunction sensor can determine deviation from the vertical. When the rotation of the earth's magnetic field is added, a three-dimensional position can be sensed.

In this way, without impeding movement, the position of the body can be analyzed in a short time using only a few markers and multifunction sensors attached to the body and analyzing the results with image editing software and a model of the human body. In the present environment that we constructed, we emphasized kinematics and left kinetics for another time. We had the caregiver stand on a force

Fig. 7.3 Markers

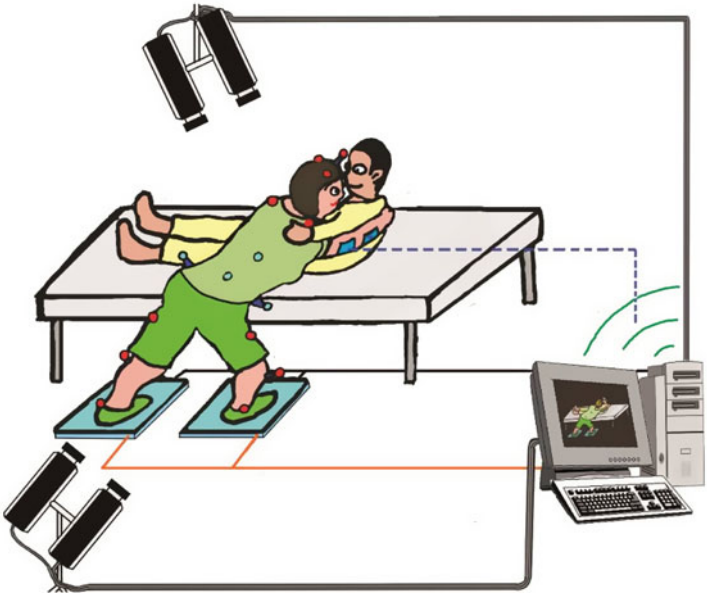


Fig. 7.4 Overview of the experiment

plate and estimated the stress put on the whole body from the force measured on the plate. In this way we analyzed the movements of a nurse or adult layperson and tested the system.

$$y = z\frac{u_H}{f}$$
$$x = z\frac{u_L}{f} - \frac{b}{2} \tag{7.1}$$

Fig. 7.5 Parallel axis cameras

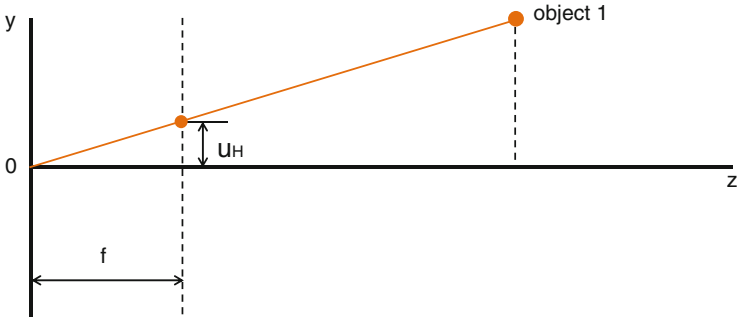


Fig. 7.6 Calculation of 3D positions (Reprinted with permission from [6]. Copyright 2013, AIP Publishing LLC)

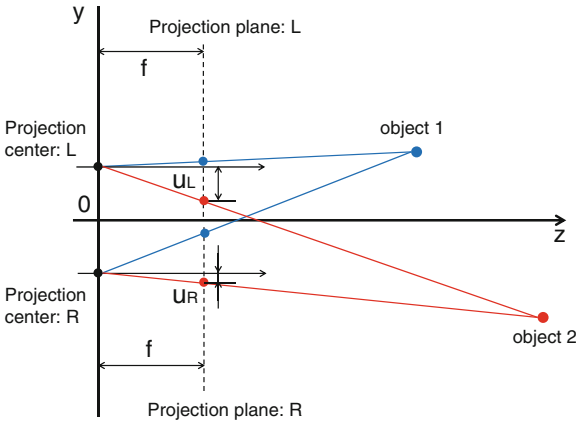


Fig. 7.7 Measuring distance using parallax (Reprinted with permission from [6]. Copyright 2013, AIP Publishing LLC)

or

$$x = z\frac{u_R}{f} + \frac{b}{2}$$

$$\begin{aligned} u_L &= f\frac{x + b/2}{z}, & u_L - u_R &= f\frac{b}{z} \\ &\Rightarrow & & \\ u_R &= f\frac{x - b/2}{z} & z &= f\frac{b}{u_L - u_R} \end{aligned} \tag{7.2}$$

3D locations were expressed as follows.

$$z = \frac{f \cdot b \cdot r}{u_L - u_R} \qquad \begin{cases} f: \text{focal distance} \\ b: \text{baseline distance} \\ r: \text{conversion ratio of pixels to mm} \end{cases} \tag{7.3}$$

Since f, b and r are constants, depth z can be calculated. If depth and parallax are known, the product of f, b and r can be calculated.

$$\begin{aligned} C &= f \cdot b \cdot r \\ depth &= \frac{C}{parallax} \end{aligned} \tag{7.4}$$

Figure 7.8 shows a multifunction sensor. These sensors output acceleration, angular speed, and the earth's magnetic field. Transmission is by Bluetooth. When an object is immobile, the angle is calculated from gravitational acceleration and orientation of the magnetic field (Fig. 7.9). Figure 7.10 shows an example of the data from the sensors.

Fig. 7.8 A multifunction sensor. This sensor is combination of tri-axial acceleration sensor and geomagnetic sensor (Reprinted with permission from [6]. Copyright 2013, AIP Publishing LLC)

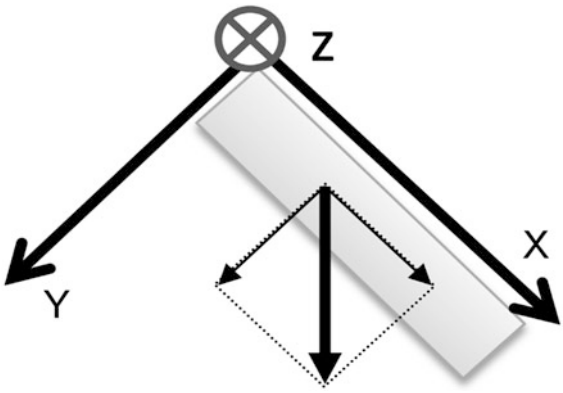


Fig. 7.9 Calculating the angle (Reprinted with permission from [6]. Copyright 2013, AIP Publishing LLC)

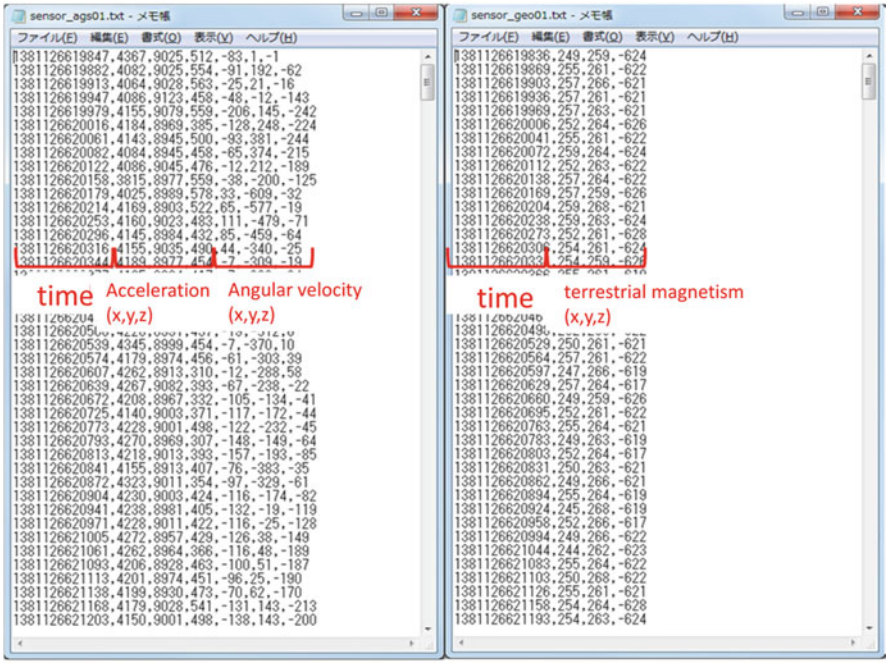


Fig. 7.10 Output data of the sensors

7.3 Detection and Analysis of the Motion of Aiding a Person to Sit Up

From the above discussions, we have constructed a prototype system and carried out experiments to confirm the appropriateness and efficacy of the proposed method.

Fig. 7.11 Aiding a patient to sit up in bed (Reprinted with permission from [6]. Copyright 2013, AIP Publishing LLC)

7.3.1 Subjects

The subjects were three healthy adults with no history of lumbago; one patient (male, aged 21, height 176 cm, weight 64 kg); two caregivers, one of whom was an adult layperson (male, aged 23, height 178 cm, weight 71.5 kg) and the other a professional nurse (female, aged 36, height 159 cm, weight 47.5 kg).

We defined "professional nurse" as a clinical nurse (including one with teaching experience) with more than 7 years of experience and who had taught aiding patients' movements on a daily basis.

Aiding a patient to sit up in bed was done as follows. The patient is lying supine in bed. The caregiver puts one upper arm under the patient's back and grasps the patient's elbow with his hand to aid the patient to sit up (Fig. 7.11).

7.3.2 Experimental Parameters

We used the same kind of bed that the hospital uses (Paramount 91). The height of this bed is adjustable. We set the height at 50 cm from the floor. To make the markers visible and to avoid glare we used black sheets. On the day before the experiment the subjects ate regular meals and had a good night's sleep.

The patient lay supine on the bed and the lay caregiver helped the patient sit up three times in whatever way he thought suitable. Next the professional nurse performed the same action three times in approved fashion. A 10-min interval was allowed between each action to avoid the danger of fatigue.

7.3.3 Measurements

For the experiments we used two pairs of CCD cameras, one personal computer, four multifunction sensors (small multifunction sensors, TSND121, ATR Pro-motions Inc.), and one 350 thousand pixel CMOS camera (DFK22 AUCO3, Argo Co. Inc.).

The multifunction sensors were attached to the non-dominant limb of the caregiver, one on the outside of the upper arm 10 cm below the shoulder, and one on the forearm 5 cm from the elbow, and their tracks were recorded during the sitting-up process. The experiments were recorded on two DVDs.

The caregivers were allowed to speak freely to the patient. At the conclusion of the experiments, the patient was queried regarding his/her physical and mental stress during the process of sitting up.

7.3.4 Results

With the present method of calculation, the posture of the entire body can be envisioned, and the following differences were noted between the professional person and the layperson.

7.3.4.1 The Manner of Using the Entire Body

The lay caregiver attempted to raise the patient using only the strength of his arm, but the professional pulled the patient toward herself as if embracing the patient and raised the patient to a sitting position. Since the lay caregiver used only his arm, the upper arm was almost perpendicular to the ground. On the other hand the professional flexed her knees and squatted, so that her shoulder, elbow, and wrist were at almost the same height (Fig. 7.12).

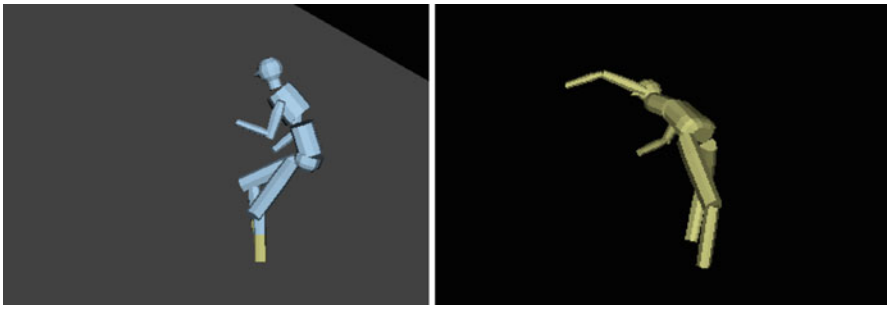


Fig. 7.12 The difference in posture of a lay caregiver and of a professional immediately after raising a patient to a sitting position. Layperson (*left*) and professional nurse (*right*) (Reprinted with permission from [6]. Copyright 2013, AIP Publishing LLC)

Table 7.1 The results of experimentally aiding a patient to sit up in bed (Reprinted with permission from [6]. Copyright 2013, AIP Publishing LLC)

		Layperson	Professional nurse
Height of shoulder, elbow and wrist above the floor		126.3/98.8/113.6 cm	105.1/96.7/103.9 cm
Angle of lumbar joint	All directions	36.7°	11.6°
immediately after raising patient	Inclined forward	36.4°	4.4°

7.3.4.2 Angle of the Lumbar Vertebrae

Table 7.1 shows the results of experimentally aiding a patient to sit up in bed. When the angle of the pelvis and the abdomen were calculated at the end of the action, it was 36.7° (36.4° when inclined forward) for the lay caregiver, but only 11.6° (4.4° when inclined forward) for the professional (Table 7.1). The lay caregiver bent the lumbar vertebrae deeply, whereas in the professional they were almost straight. The body of the lay caregiver was inclined deeply forward, exerting a load on the lumbar vertebrae.

7.3.4.3 The Patient's Impressions of the Caregivers

The patient was interviewed at the end of the experiments, with the following results. The lay caregivers were given no instructions, but had to just follow their own instincts, relying only on strength, so he found that worrisome. The professional acted after being given sufficient instructions and after having received the patient's consent. The patient sat up without effort, or as if he had not received aid at all. He found no cause for worry regarding the professional.

7.3.5 Discussion

Using our newly developed system, we analyzed the actions of a layperson and a professional nurse, and evaluated their burdens. Since these experiments placed emphasis on kinematics, difference between the postures of the professional person and the layperson could be noted. In other words, the professional adopted a posture that did not place stress on the lumbar vertebrae, but the layperson adopted a dangerous posture. The patient also was aware of the difference between the two types of caregivers, and experienced both physical and psychological anxiety regarding the layperson.

7.4 Evaluation (Kinetic Analysis) of the Burden of Nursing Care in Relation to the Use of a Force Plate and a Human Model

In the previous section we analyzed only the kinematics of the motion of the human body. In this section we will discuss the kinetics of the physical burden on the caregiver's body.

The National Institute for Occupational Safety and Health (NIOSH) sets the permissible force bearing on the lumbar vertebrae during work at 3,400 Newtons (N). We measured experimentally the lumbar burden, and for its evaluation we measured the force brought to bear on the lumbar vertebrae and the shearing force which is the cause of lumbago.

7.4.1 Purpose

To measure the burden on the lumbar vertebrae when a professional nurse and a layperson attempt to aid a patient to sit up in bed.

7.4.2 Methods

7.4.2.1 Subjects

A professional nurse without a history of lumbago (age 60, height 158 cm, weight 58 kg), and one layperson without a history of lumbago (age 22, height 176.0 cm, weight 69.0 kg); and one patient (age 22, height 174.0 cm, weight 61.0 kg).

7.4.2.2 Ethical Considerations

The potential subjects were given orally an explanation about the purpose of the experiments, how they were to be performed, how the subjects were free to participate, and a promise of confidentiality. Those who agreed to participate did so in writing and received in writing an explanation of the purpose, significance, methods, safety, and promise of confidentiality.

7.4.2.3 Experimental Methods

LED markers were attached to the caregivers. The markers were attached to points on the surface of the body where the underlying bones could be palpated.

Fig. 7.13 Experiment scene (*upper photos*: a layperson, *lower photos*: a professional nurse)

Each caregiver stood on force plates (Kisllar 9286A) on the floor and attempted to raise the supine patient into a sitting position. Each foot of the caregiver pressed on a separate plate. The markers were photographed in 3 dimensions from 2 angles with 2 parallel axis cameras. To remove influences of magnetic field sensing by multifunction sensors due to metal frame, we used the wooden bed. The wooden bed was set at a height of 40 cm from the floor (Fig. 7.13).

7.4.2.4 Measuring the Moment of the Lumbar Joints and the Shearing Stress Brought to Bear on the Lumbar Vertebrae

We calculated the moment of the lumbar joints from the inverse dynamics of the position of the markers and the pressure on the force plates. Inverse dynamics form a pair with forward dynamics. Forward dynamics calculates the movement (change of position, speed, acceleration, and angle of joints) of a body from the parallel forces and torque acting on a body. Inverse dynamics calculates whether a body has moved at a given time point from the parallel forces and torque that are presumed to be acting on a body.

The pain of lumbago is said to arise between lumbar vertebrae 4 and 5, or between 5 and the sacral vertebrae. In the present experiments we chose the moment of lumbar vertebra 5 as the center of rotation.

In the present experiments we used inverse dynamics with the Newton-Euler method to calculate the stress imposed on the lumbar region.

7.4.3 Results

The shearing stress and the pressure on the lumbar vertebrae and shown in Figs. 7.14 and 7.15. In the layperson immediately after raising the patient, shearing stress jumped to above 350 N and peaked at 480 N. It was only half of this (150 ∼ 240 N) in the professional nurse (Figs. 7.14 and 7.15). Compressive force in the professional

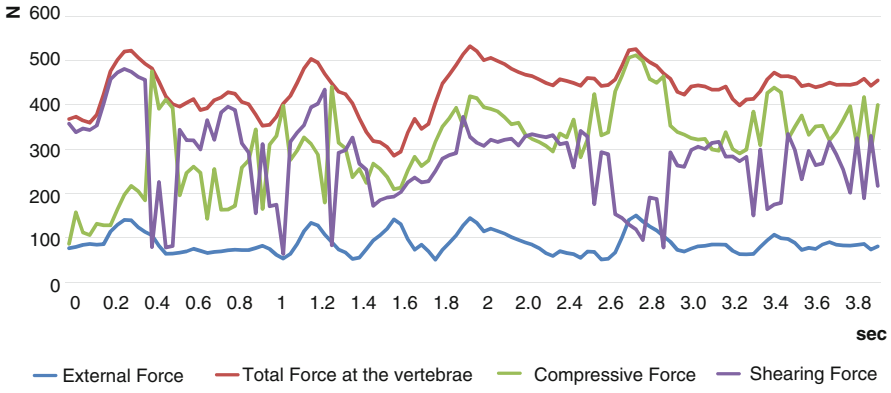


Fig. 7.14 The compressive force and the shearing force of the layperson during the experiment

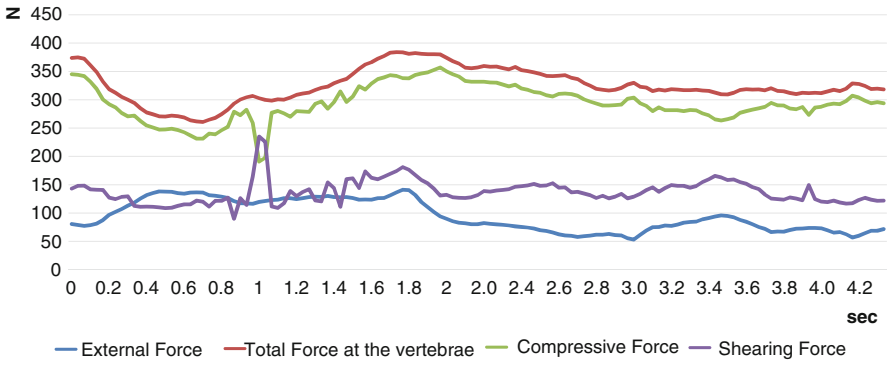


Fig. 7.15 The compressive force and the shearing force of the professional nurse during the experiment

nurses remained steady at 190 $\sim$ 350 N from the beginning to completion of the action, but fluctuated wildly from 100 $\sim$ 500 N in the layperson.

7.4.4 Discussion

In the nurse compressive force was high, but shearing force was weak. Consequently there was little danger of getting lumbago from raising a patient to a sitting position. On the other hand, shearing forces were strong in the layperson and compressive force was weak, stressing the lumbar region and causing shearing at the joints. This may be due to the differences in the way body mechanics is taught in nursing school. Even experienced nurses do not always perform actions protective of the lumbar region. This may be why lumbago is of frequent occurrence in nurse. With our analysis of the kinematics of the lumbar vertebrae angles in the present experiments, we have shown the difference between beginners and experienced nurse. Together with our data on kinetics, steps can be taken to reduce the frequency of lumbago in nurse and to assure the comfort of patients.

Concluding Remarks and Future Problems

The present research was carried out to analyze the actions of caregivers when aiding a patient to sit up in bed. The new system showed that three dimensional analysis could be performed even regarding points on the subjects' bodies that were hidden from view. Using this system we were able to evaluate the performance of both lay and professional caregivers. We found a clear difference between the performances of the two types of caregivers, and noted that the professional adopted a posture that was safe and did not stress the lumbar vertebrae, whereas the layperson adopted an unsafe posture.

Acknowledgements This research was supported by a grant (Basic Research C, Project No. 1224593219; Principal Investigator, Akemi Nomura) from the Ministry of Education and Science, Japan.

References

1. D. Daynard, A. Yassi, J.E. Cooper, R. Tate, R. Norman, R. Wells, Biomechanical analysis of peak and cumulative spinal loads during simulated patient-handling activities: a substudy of a randomized controlled trial to prevent lift and transfer injury of health care workers. App. Ergon. **32**(3), 199–214 (2001)
2. M. Inada, S. Masuda, E. Futatsumori, Analysis of the performance of student nurses when moving a patient to a wheelchair (in Japanese). Bull. Jpn. Red Cross Coll. Nurs. **13**, 43–50 (1999)

3. K. Itami, Y. Yasuda, K. Otsuki, M. Ishibashi, T. Maesako, Developing a body-mechanistic checking system based on forward inclination when bedmaking (in Japanese). Jpn. J. Educ. Technol. **33**(1), 1–9 (2009)
4. R. Jang, W. Karwowski, P.M. Quesada, D. Rodrick, B. Sherehiy, S.N. Cronin, J.K. Layer, Biomechanical evaluation of nursing tasks in a hospital setting. Ergonomics **50**(11), 1835–1855 (2007)
5. Y. Mito, Dynamic analysis of caregiver foot position when moving a hemiplegic patient to a wheelchair – a fundamental study (in Japanese). J. Kanagawa Univ. Hum. Serv. **1**(1), 7–18 (2004)
6. A. Nomura, Y. Ando, T. Yano, Y. Takami, S. Itou, T. Sato, A. Nemoto, H. Arisawa, Development of a motion capture system for measuring hidden points on a human model and its application to aiding a patient to sit upright in bed, in *International Symposium on Computational Models for Life Sciences*, Sydney, ed. by C. Sun, T. Bednarz, T.D. Pham, P. Vallotton, D. Wang. AIP Conference Proceedings, 2013, pp. 221–229
7. T. Okuda, K. Kodama, T. Mineshima, K. Iriuchijima, O. Nitta, T. Tsutsui, M. Sakamoto, An experimental study of spatial conditions when shifting a patient (in Japanese). Rigakuryoho Kagaku **11**(2), 63–70 (1999)
8. B. Schibye, A.F. Hansen, C.T. Hye-Kundsen, M. Essendrop, M. Bodher, J. Skotte, Biomechanical analysis of the effect of changing patient-handling technique. Appl. Ergon. **34**(2), 115–123 (2003)
9. J.H. Skotte, M. Essendrop, A.F. Hansen, B. Schibye, A dynamic 3D biomechanical evaluation of the load on the low back during different patient-handling tasks. J. Biomech. **50**(10), 1357–1366 (2002)
10. E.R. Vieira, S. Kumar, Safety analysis of patient transfers and handling tasks. Qual. Saf. Health Care **18**, 380–384 (2009)

Chapter 8
Classifying Epileptic EEG Signals with Delay Permutation Entropy and Multi-scale K-Means

Guohun Zhu, Yan Li, Peng (Paul) Wen, and Shuaifang Wang

Abstract Most epileptic EEG classification algorithms are supervised and require large training datasets, that hinder their use in real time applications. This chapter proposes an unsupervised Multi-Scale K-means (MSK-means) algorithm to distinguish epileptic EEG signals and identify epileptic zones. The random initialization of the K-means algorithm can lead to wrong clusters. Based on the characteristics of EEGs, the MSK-means algorithm initializes the coarse-scale centroid of a cluster with a suitable scale factor. In this chapter, the MSK-means algorithm is proved theoretically superior to the K-means algorithm on efficiency. In addition, three classifiers: the K-means, MSK-means and support vector machine (SVM), are used to identify seizure and localize epileptogenic zone using delay permutation entropy features. The experimental results demonstrate that identifying seizure with the MSK-means algorithm and delay permutation entropy achieves 4.7 % higher accuracy than that of K-means, and 0.7 % higher accuracy than that of the SVM.

Keywords Unsupervised learning • Delay permutation entropy • MSK-means • SVM • Seizure detection • Epileptogenic focus location

8.1 Introduction

Epilepsy is a prevalent neurological disorder stemming from temporary abnormal discharges of the brain electrical activities and leading to unprovoked seizures. About 1 % population in the world are diagnosed as epilepsy [19]. Fortunately,

G. Zhu (✉)
Faculty of Health, Engineering and Sciences, University of Southern Queensland, Toowoomba, QLD 4350, Australia

School of Electronic Engineering and Automation, Guilin University of Electronic Technology, Guilin, Guangxi 541004, China
e-mail: guohun.zhu@usq.edu.au

Y. Li • P. (Paul) Wen • S. Wang
Faculty of Health, Engineering and Sciences, University of Southern Queensland, Toowoomba, QLD 4350, Australia
e-mail: yan.li@usq.edu.au; peng.wen@usq.edu.au; shuaifang.wang@usq.edu.au

C. Sun et al. (eds.), *Signal and Image Analysis for Biomedical and Life Sciences*, Advances in Experimental Medicine and Biology 823,
DOI 10.1007/978-3-319-10984-8_8

EEG recordings can show the brain electrical activity information and provide valuable insight into disorders of the brain. EEG signals are considered as important data in diagnosing epilepsy and predicting epilepsy seizures. However, the traditional visual inspection by analysts is time-consuming, error prone and not sufficient enough for reliable detection and prediction. The randomization nature of epilepsy seizures and their large EEG recording datasets make epileptic EEG classification more difficult. Hence, an automatic epileptic classification system is becoming more and more on demand.

Epileptogenic zone (EZ) is a cortical area indispensable for the generation of seizures. Resection of the EZ is the most direct and effective means to drug-resistant patients [10]. Identifying the EZ is regarded as a prerequisite for successful surgical treatment [24]. Epileptologists analyse features from video EEG, intracranial EEG (iEEG), magnetoencephalography [31], positron emission tomography or single-photon emission signals [15] to detect which areas of the cortex induces seizures during ictal state. Among them, the iEEG signals recorded from surface of the brain are widely used by experienced epileptologists. The EZ has high interelectrode template similarity based on frequency entropy template from 96 electrodes iEEG [7]. Time-variant connectivity analysis method was used to localize the ictal-onset zone from iEEG recordings [29]. Frequency domain source imaging approach was applied to identify epileptogenic zones in patients with secondary generalized epilepsy [8]. Because seizures often occur randomly and patients look and feel normal during seizure-free intervals, it is possible to have an extended normal period (5–10 days) before catching a seizure. The time-consuming process also increases the risk of infection.

Most of traditional automatic epileptic classification systems use supervised learning classifiers, such as artificial neural networks (ANN), support vector machines (SVMs) and decision trees. Guo et al. [11] applied wavelet discrete transform features and an ANN for discriminating ictal EEGs from normal EEGs. Nicolaou and Georgiou [20] employed permutation entropies (PE) and SVMs to determine the seizure from EEG signals. Subasi [28] fed wavelet features to a fuzzy classifier to identify seizure from EEG signals. Siuly et al. [25] proposed a clustering technique to classify ictal and healthy EEGs. Song and Liò [26] classified ictal, inter-ictal and normal EEGs by features based on sample entropy (SE) and an extreme learning machine algorithm. Zhu et al. [32] implemented visibility graph (VG) based features and a discriminant classifier to identify ictal EEGs from healthy EEGs. However, an automatic epileptic classification system normally requires large sets of data to train a classifier, and to improve the accuracy. Meanwhile, all the data are normally required in a specific format to meet certain conditions. For example, the number of data segments/epochs should be the same in the training data and testing data. Besides, the target categories for all the data segments in the training set rely on the labels obtained manually by experts. All these limitations impede the current supervised epileptic EEG classification techniques from being used.

K-means clustering is a popular unsupervised learning method which was first presented by MacQueen [17]. It consists of two simple steps: the first step is to randomly choose k centroids for k clusters. The second step is to separate the

input data into k disjoint clusters according to the distance between each data instance and the k chosen centroids. Its simplicity and fast clustering make it easy to implement. However, if some data points belonging to the same cluster are incorrectly assigned into other disjoint clusters during the first step, it may lead to wrong classification results. Recently, Vattani [30] showed that the running time of the K-means algorithm increases exponentially when the data size increases. To solve the cluster initialization issue, Arthur and Vassilvitskii [4] proposed a K-means++ algorithm and improved the classification accuracy by initializing the centroids one by one. Bahmani et al. [5] reported that the K-means++ did not work well on large sets of data because it relies too much on the central point initialization. Liu et al. [16] presented an improved radius K-means algorithm to improve the accuracy for clustering network datasets.

The delay permutation entropy (DPE) is proposed to extract features from single channel EEG signals in this chapter. The optimal delay factor is obtained from comparing the performance from zero to 50 delay lags. The extracted DPE features are then forwarded into a multi-scale K-means (MSK-means) classifier to discriminate epileptic EEGs from healthy EEGs and to identity the epileptogenic zone signals from non-epileptogenic zone signals. The experimental results showed that the DPE features with MSK-means algorithms could distinguish epileptic EEGs and localize epileptic zone effectively.

This chapter is organized as follows: In Sect. 8.2, the experimental dataset is introduced. The traditional K-means algorithm and the proposed MSK-means method are described in Sect. 8.3. In Sect. 8.4, the comparison results of the K-means, MSK-means and SVM with the DPE features to identify epileptic EEG are presented. Classifying EEG signals from epileptogenic zone with DPE from non-epileptogenic area are shown in Sect. 8.5.

8.2 Experimental Data

There are two databases used in this chapter. The first one was described by Andrzejak et al. [1], which was digitized at 173.61 samples per second from a 12-bit A/D convertor. The band-pass filter setting was 0.53–40 Hz. The whole database is made up of five EEG datasets (denoted as sets A-E), each containing 100 single-channel EEG signals from five separate classes and 4,097 data points. Sets A and B were recorded from five healthy volunteers with eyes opened and eyes closed, respectively. Sets C and D were recorded from the EEGs of epileptic patients during seizure-free intervals from the opposite hemisphere of the brain and within the epileptogenic zone, respectively. Set E contains the seizure activity EEGs.

The second were obtained from a public Bern-Barcelona EEG database [3] collected from ten patients. The database includes two distinct sets: one comes from epileptogenic zone (set F) and the other is recorded from brain areas that were not involved in seizure onset (set N). The sample rate is 512 Hz if the number of record is less than 64 channels. Otherwise, it is 1,024 Hz. Each data contains two signals,

signal x is the patient's focal EEG channel and signal y is the neighboring channel of the epileptogenic zone. Each signal in each recording has 10,240 data points. The detailed description and usage of this database can be found in [2]. This study divides the Bern-Barcelona EEG database into two sub datasets. The number of recordings in these sets are 100 and 4,500 respectively. The small set is named #50 which is the sample data used by [2], and the large set is denoted as #750.

8.3 Methodology

The proposed epileptic classification system is shown in Fig. 8.1. The extracted features based on delay permutation entropy from the raw EEG data are directly transferred to a MSK-means classifier for epilepsy classification. The K-means clustering algorithm and the SVM classifier in Fig. 8.1 are used for comparison purpose.

8.3.1 K-Means Algorithm and K-Means++ Algorithm

Given a set of observations $X = (x_1, x_2, \ldots, x_n)$, the K-means clustering technique aims to partition n observations into k sets ($k \leq n$) $C = (c_1, c_2, \ldots, c_k)$ based on the Euclidean distance. The Euclidean distance between the ith data point and the jth centroid is defined as follows:

$$d(x_i, c_j) = \sqrt{\sum_{j=1}^{k} (x_i - c_j)^2} \tag{8.1}$$

The central point of a cluster is recomputed as:

$$C_j = \frac{1}{|C_j|} \sum_{x \in c_j} x \tag{8.2}$$

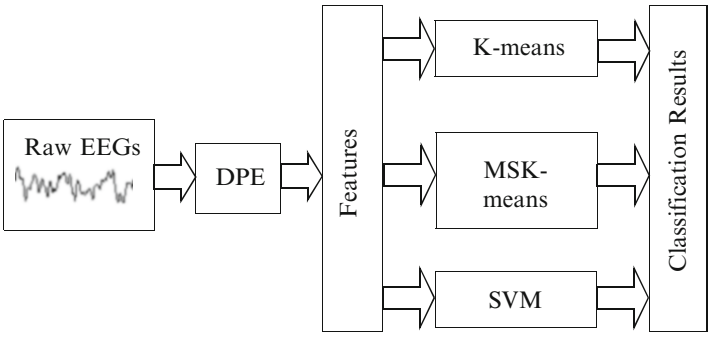


Fig. 8.1 The structure of the proposed epileptic EEGs classification system

The K-means algorithm minimizes the within-cluster sum of squares by Lloyd iteration to make the data of the same cluster more compact and dependent:

$$\omega = \sum_{j=1}^{k} \sum_{i=1}^{|C_j|} d(x_j, c_i) \tag{8.3}$$

The main idea of the K-means algorithm is to randomly choose k observations as the cluster central points (centroids) and assign all the remaining data to their nearest centroids based on Eq. (8.1). Then the new centroid of each cluster is calculated using Eq. (8.2). The algorithm converges when the new centroids are same as the old centroids. The randomness of initialization is error prone if some data points from the same class are assigned to different cluster centroids. The K-mean++ algorithm proposed by Arthur and Vassilvitskii [4] improves the initialization by the following algorithm:

```
Algorithm 1.  K-means++ algorithm init
Input: X, k
N=number of X
C=randomly choose a point from X
While |C|<k begin
            Dist [1...n]=the distance between X and C
            U=sum(Dist[1...n])
            J=1
            Do begin U=U-Dist[j], j=j+1 end while U>0
            C=C union X[j]
end while
```

The K-means++ algorithm needs additional computation time for initializing centroids. However, the time complexity of both K-means and K-means++ algorithms are NP-hard in [4].

8.3.2 Multi-scale K-Means (MSK-Means) Algorithm

The scale of initialization of both K-means and K-means++ is small and limited to the data size, which is not suitable for large sizes of EEG signals. In this chapter, an MSK-means algorithm is proposed to improve the performance by optimizing the cluster initialization.

The concept of multi scale analysis of time series was first proposed by Costa et al. [9]. The multi scale technique transfers one dimensional time series $\{x_t\}_{t=1,\ldots,n}$ into another time series $\{y_t\}_{t=1,\ldots,\frac{n}{\tau}}$ with a different scale. Here τ is a scale factor. The transformation formula is as follows:

$$y = \frac{1}{\tau} \sum_{i=(j-1)\tau+1}^{j\tau} x_i, \ 1 \le j \le \frac{n}{\tau} \tag{8.4}$$

Based on Eq. (8.4), the original algorithm is adjusted as:

```
Algorithm 2.  MSK-means init
Input: X, k, and scale factor
 Y construct according to above eq. and scale factor
 Med=K median positions of Y
 C=empty set
 I=1
 While (i<k)
        C[i] = random a point in Med[i:i+1]
 end
```

Similar to the K-means++ algorithm, the MSK-means algorithm only improves the initialization part of the K-means algorithm. Lloyd repeat is conducted with the scaled time series Y instead of the original times series X. The computational complexity of the MSK-means algorithm is as follows.

Theorem 8.1 ([34]) *Let us assume that n is the number of the datasets, d is the number of iterations, k is the number of clusters and τ is a parameter, the time complexity of the MSK-means algorithm is $O(\max\{\frac{ndk}{\tau}, n\})$.*

Proof In the MSK-means algorithm, the time complexity of Eq. (8.4) is n. It indicates that the complexity of k median value is $\frac{n}{\tau}$. The time complexity of Lloyd repeat is $O(\frac{ndk}{\tau})$. The time complexity of the MSK-means algorithm is $O(\max\{\frac{ndk}{\tau}, n\})$. □

According to Theorem 8.1, the time complexity of the MSK-means algorithm can be $O(n)$ when τ is large enough, which means it can be of higher efficiency than both the K-means and K-means++ algorithms. The relation of τ and the time complexity of the multi-scale K-means algorithm is discussed in Sect. 8.4.

8.3.3 Delay Permutation Entropy

In general, EEG time series is nonlinear and complicated. The amplifies of the two EEG time series are possibly different even for a same test subject when these two signals are sampled at different times. To reduce the disturbances of white noise, permutation entropy [6] is used for the permutation symbols to substitute the raw EEG data points to calculate the entropy. The DPE improves the nonlinear detection performance of PE for moving-average processing time series by Matilla-García [18]. Given a time series $\{x_i\}_{i=1,2,\ldots,n}$, the DPE algorithm is outlined as follows [18].

1. Convert the time series X into an m dimensions sequence of vectors X_m using time delay λ embedding by:

$$X_m\{t\} = \{x_t, x_{t+\lambda}, \ldots, x_{t+(m-1)\lambda}\} \; ; \; 1 \le t \le (n - m\lambda + 1) \tag{8.5}$$

where $2 \le m < n$.

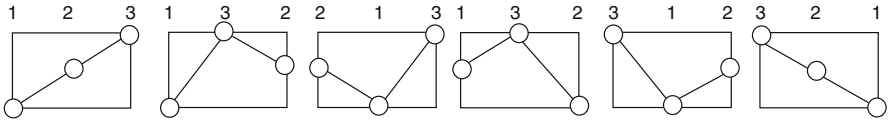


Fig. 8.2 Six possible ordinal patters for m=3

2. Let $j_k \in \{t, t+\lambda, \ldots, t+(m-1)\lambda\}$, and then arrange each vector of X_m to an increasing order.

$$X'_m\{t\} = \{x_{t+j_1}, x_{t+j_2}, \ldots, x_{t+j_m}\} \; ; \; 1 \le t \le (n - m\lambda + 1) \tag{8.6}$$

Let $\pi = \{j_1, j_2, \ldots, j_m\}$, then π_i is a symbol of $m!$ possible permutations of a vector $\{1, 2, \ldots, m\}$. As shown in Fig. 8.2, for $m = 3$, there are six possible ordinal patterns between $X_3(t) = \{x_t, x_{t+\lambda}, x_{t+2\lambda}\}$.

For example, if $X_3(t) = \{6, 3, 11\}$, then $\pi_3 = \{2, 0, 1\}$.

3. From $i = 1$ to $n - m\lambda + 1$ calculate each $X_m\{t\}$ into a symbol π_i.
4. Let $p(\pi_i, \lambda)$ be the probability distribution of π_i.
5. Calculate the Shannon entropy of all the symbols

$$h(X, m, \lambda) = -\sum_{j=1}^{n-m\lambda+1} p(\pi, \lambda) \ln(\pi, \lambda) \tag{8.7}$$

Eq. (8.7) is derived from Shannon entropy, so it reflects the regularity of the time series $\{x_i\}_{i=1,2,\ldots,n}$. The values of DPE are $\le h(X, m, \lambda) \le \ln(m!)$ [18].

The smaller $h(X, m, \lambda)$ is, the more regular the time series are, such as an increment, decrement or periodicity. When $h(X, m, \lambda)$ is zero, the time series is a constant value. On the contrary, the bigger $h(X, m, \lambda)$ is, the time series are in a more random order. All the values of long enough time series are independent to each other without any rules to follow when $h(X, m, \lambda)$ is close to the ceiling $\ln(m!)$.

In order to compare different values of DPE associated with different m, one of the method is to normalize the DPE with a factor $m - 1$. The permutation entropy of each symbol is defined as follows:

$$h(X, m, \lambda) = \frac{-1}{m-1} \sum_{j=1}^{n-m\lambda+1} p(\pi, \lambda) \ln(\pi, \lambda) \tag{8.8}$$

There is also another normalization method, which is demonstrated as follows:

$$h(X, m, \lambda) = \frac{-1}{\ln(m!)} \sum_{j=1}^{n-m\lambda+1} p(\pi, \lambda) \ln(\pi, \lambda) \tag{8.9}$$

Eq. (8.9) is always less than or equals to one.

8.3.4 Detecting Nonlinear Structure with DPE

The minimum of $h(X, m, \lambda)$ is zero when the input signals is a constant. However, even when signals appear to be not so regular, particular value of λ could make the value of $h(X, m, \lambda)$ as zero. For example, given $X = (8, 4, 1, 9, 6, 3, 11, 7, 6)$, then $h(X, 2, 1)$ is 0.24, when $\lambda = 3$, $h(X, 2, \lambda)$ becomes zero and $h(X, 2, 4)$ is 0.29. Therefore, the lag $\lambda = 3$ represents the fixed delay factor of a given signal.

Matilla-García and Ruiz proposed a theorem as follows.

Theorem 8.2 ([18]) *Let $\{x_i\}_{i=1,2,\ldots,n}$ be the stationary time series and $h(X, m, \lambda)$ is defined in Eq. (8.7) for a fixed embedding dimension $m > 2$, with $m \in N$. If the most relevant lag of the time series $\{x_i\}_{i=1,2,\ldots,n}$ is λ_0 then the following equation holds*

$$h(X, m, \lambda_0) \leq \min(h(X, m, \lambda)) \tag{8.10}$$

In this study, we extend Theorem 8.2 to the following lemma.

Lemma 8.1 *$h(X, m, \lambda_0) \leq \min(h(X, m, \lambda))$ holds when $m \geq 2$ and $\lambda < n-m-1$.*

According to the PE algorithm in [6], PE is a special case of DPE when $\lambda = 1$. During this study, it assigns $m = 4$ to identify the seizure and $m = 3$ to recognize EZ.

8.3.4.1 Support Vector Machine

To compare the performance of the unsupervised MSK-means algorithm with the supervised classifiers, the support vector machine (SVM) is selected to conduct the binary classification. The SVM has been successfully used in epileptic EEG classification [20,34]. It can perform both linear space discrimination and nonlinear classification by choosing different "kernel" functions which can be linear, polynomial kernel, radial basis function (RBF) and sigmoid. In this chapter, the SVM algorithm with RBF kernel is used from the R package e1071 [13].

8.4 Experiments for Detecting Seizures

To evaluate the performance of the MSK-means algorithm presented in Sect. 8.3, C programming language is used, while the SVM and K-means algorithms are implemented by the R package e10171 and stats package, respectively. The experiments include three parts: (1) evaluating the delay lag λ and DPE indices based on different groups of EEG; (2) evaluating the accuracy for classifying epileptic EEGs and healthy EEGs with different λ and DPEs; and (3) comparing the accuracy level of the K-means, SVM and MSK-means for identifying seizures. For experiments

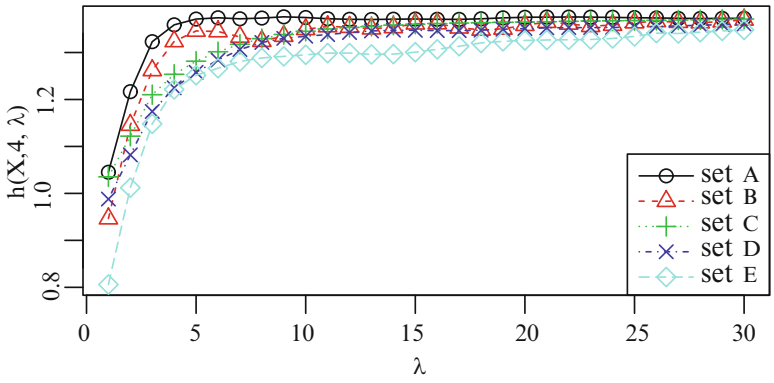


Fig. 8.3 The relation between delay lag λ and the DPE indices for $m = 7$ for five groups of EEG signals

(1) and (2), the DPE indices are changed from 1 to 30. During these experiments, each EEG recording is divided into four equal epochs, and a total of 4,000 new EEG segments are produced. During the SVM classification processing, the extracted features of odd EEG segments are used in the training dataset while those of even epochs are used in the testing dataset.

8.4.1 Evaluating Delay Lag for Exact Order m

This section serves to evaluate DPE of the delay lag λ associated with five EEG groups. Figure 8.3 shows the relation between delay lag λ and the DPE indices.

Based on Fig. 8.3, the DPE indices associated with epileptic EEG are close when delay lag λ is larger than 20. The healthy EEGs (sets A and B) become horizontal shape after $\lambda > 5$. Our findings agree with the results from Popov and Avilov [23]: Comparing DPE indices from healthy EEGs with those associated with epileptic EEG, the epileptic EEG signals are more regular than healthy because DPE indices on healthy is large from $\lambda = 1$ to $\lambda = 30$.

8.4.2 Accuracy of Detecting Seizures with Different λ DPE

In this section, we use the K-means, MSK-means and SVM algorithms to discriminate seizure EEGs (set E) from interictal EEGs (sets C and D). The results are demonstrated in Fig. 8.4. In order to obtain good performances, the values of scale factor τ of MSK-means are selected as eight; and λ changed from 1 to 30.

Figure 8.4 shows that the accuracy of the MSK-means algorithm with $\tau = 8$ archives 100.0 % when $\lambda = 1$. From $\lambda = 7$ to $\lambda = 25$, the accuracies with MSK-means is better than K-means and SVM. The peak of accuracy of SVM is 92.0 %

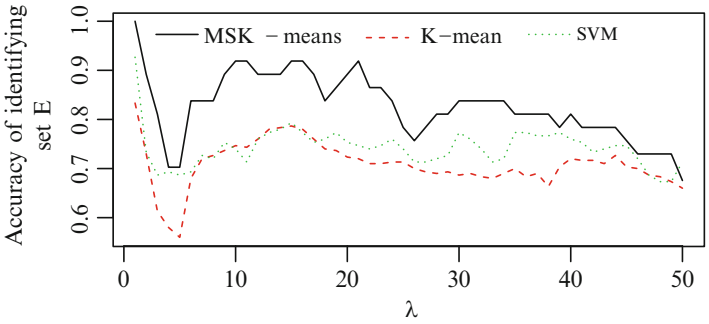


Fig. 8.4 Identifying seizure EEG (set E) and inter-ictal EEG (sets C and D) with different λ DPE and three classifiers

Table 8.1 The average 10-times classification accuracy and execution time of the three algorithms with different pairs of datasets ($\tau = 10$, $n = 1{,}024$) with DPE features

	SVM		K-means		MSK-means	
Data groups (set)	Accuracy (%)	Time (ms)	Accuracy (%)	Time (ms)	Accuracy (%)	Time (ms)
A vs. E	99.3	30	95.3	20	100.0	20
B vs. E	95.0	35	79.3	20	90.0	20
C vs. E	97.5	35	90.3	20	100.0	20
D vs. E	93.5	40	76.5	25	92.5	20
(A,B,C,D) vs. E	95.7	70	81.6	35	90.0	35

when $\lambda = 1$. Because accuracies just use only one dimension features, which imply that the healthy EEG and epileptic EEG could be effectively distinguished by DPE with a fixed delay lag.

8.4.3 Identifying Seizures from Other Sets

In this section, the performances of the K-means, SVM and MSK-means algorithms for different pairs of datasets are presented. Four same size of datasets containing 1,024 epochs and the seven DPE indices ($\lambda = 1$–7) extracted features from each epoch are used. The results are shown in Table 8.1.

From Table 8.1, the classification accuracy for the pair of A vs. E can archive 100 % using the DPE indices and MSK-means classifier, while the performance of pair (A,B,C,D) vs. E is minimally improved over that of SVM classifier. However, it can be further improved by changing the τ selection (e.g., it achieves 96 % accuracy when $\tau = 40$).

The classification accuracies on the epileptic EEG database from different literature are presented in Table 8.2.

Table 8.2 The classification accuracy by the MSK-means and other existing methods

Researchers	Features (epochs length)	Classifier	Datasets	Accuracy (%)
Polat and Güneş [22]	PSD ($n = 256$)	Decision Tree	A,E	98.7
Guo et al. [11]	DWT ($n = 4{,}097$)	ANN	A, E	99.6
Nicolaou and Georgiou [20]	PE ($n = 1{,}024$)	SVM	A, E	93.4
			C, E	88.8
Zhu et al. [34]	Entropy (n=1,024)	MSK-means	A, E	100.0
[a]Orhan et al. [21]	DWT and K-means ($n = 4{,}097$)	ANN	A, E	100.0
			(A,B,C,D), E	99.6
The proposal method	DPE ($n = 1{,}024$)	MSK-means	A, E	100.0
			C, E	100.0
			(A, B,C,D), E	96.0

[a]The K-means algorithm was used by Orhan et al. [21] as a feature extraction method instead of a classifier

Based on Tables 8.1 and 8.2, the proposed MSK-means method has better performance in distinguishing the epileptic EEGs from healthy EEGs, especially in identifying the epileptic EEGs from normal EEGs. Without clinical history data records, it is impossible for a supervised algorithm to conduct classifications, while the MSK-means algorithm can work well because it is unsupervised.

8.5 Experiments for Detecting Epileptic Zone

To evaluate the performance of the methods in Sect. 8.3, the DPE features are analysed on epileptogenic zone and nonepileptogenic zone. The experiments include two parts: (1) analysing DPE index associated with epileptic iEEGs under different delay lags; (2) evaluating classification accuracy of the DPE features by different lags on two different sizes of datasets. Both experiments investigate the datasets #50 and #750 and the value of λ ranges from 1 to 50.

8.5.1 Statistical Analysis of Relation Between DPE Index and λ

Figure 8.5 shows the relation between different λ and the DPE index associated with channel x of epileptogenic zone iEEG and non-epileptogenic area iEEGs on two datasets. In order to evaluate the relation between the DPE index and λ, the value of λ changes from 1 to 50.

To make the relation between DPE index and λ more clear, comparison between DPE index of epileptogenic iEEG and those of non-epileptogenic signal are employed with the Student's test. Two ranges of λ are measured based on dataset

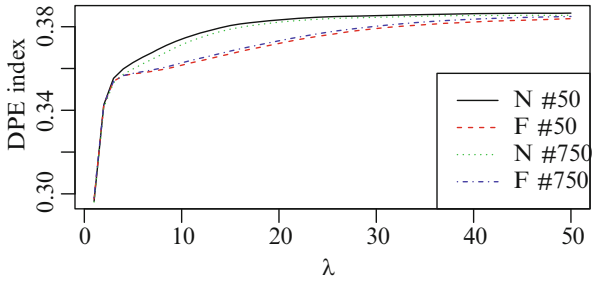


Fig. 8.5 The relation between the DPE index and λ on iEEG channel x

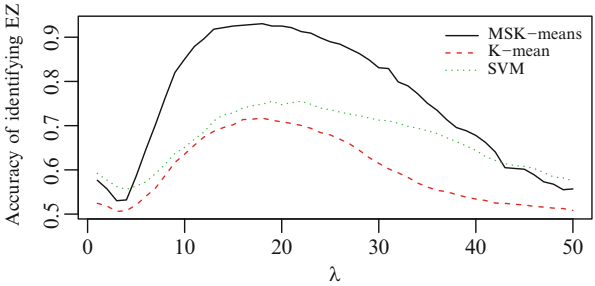


Fig. 8.6 The relation between accuracy and λ of DPE on iEEG channel x for the identification of epileptogenic and nonepilptogenic

#750. The first range is from 1 to 30, the statistical DPE indices between sets F and N are not significantly different ($p = 0.08$). The second range is from 5 to 30, and the statistical difference are considered significant ($p = 0.01$).

8.5.2 Relation Between λ and Classification Accuracy

This section investigates the relation between classification accuracy and λ when one dimension DPE index is applied to identify the epileptogenic iEEG. Firstly, each recorded signal x is extracted with DPE indices when λ changing from 1 to 50. For set #50, a total of 100 by 50 dimensional DPE features are extracted, and for set #750, a total of 4,500 by 50 dimensional DPE indices are also extracted. Each dimensional feature is forwarded into a K-mean, MSK-means and SVM to conduct classification, where the odd indices of instances are used for training and the remaining features are used for testing. Finally, the relation of the accuracies and λ on two EEG datasets are obtained and illustrated in Fig. 8.6.

Based on Fig. 8.6, the maximum accuracy is 93 % for the EEG set #750 and λ is selected as 18 when MSK-means classifier ($\tau = 8$) is applied. It is clear that the accuracies of MSK-means is better than SVM and K-means when λ is located

Table 8.3 Classification accuracy with PE features and SVM (half for training and others for testing)

	Set F	Set N
Set F	642	435
Set N	483	690

Table 8.4 Classification accuracy with DPE index ($\lambda = 21$) and MSK-means ($\tau = 8$)

	Set F	Set N
Set F	269	13
Set N	26	254

between 6 and 30. To be more clear, a confusion matrix between the experts scoring and SVM based on PE ($\lambda = 1$ of DPE) is shown in Table 8.3. The accuracy is 59 %.

The confusion matrix based on the DPE index ($\lambda = 18$) with MSK-means classifier is shown in Table 8.4.

The accuracy in Table 8.4 is 93 %. According to Tables 8.3 and 8.4, the classification accuracy based on DPE index is 34 % higher than those based on PE and SVM. Compared to previous work [33], which is just 84 % the maximum accuracy based on SVM classifier with 50 tested recordings, this result is higher accuracy. Four thousand and five hundred testing recording shows that the result is more robust. More importantly, the highest accuracy based on the DPE index in this study is higher than the existing recorded results, which is 50–80 % [12, 14, 27]. This proposed method exhibits that DPE indices and MSK-means classifier can be potentially applied in epileptogenic focus location based on the iEEG signals.

Conclusions

Unsupervised classification algorithms play an important role in epilepsy detection. The proposed MSK-means algorithm in this study optimizes the initialization stage to improve the classification performance. Both theory and experimental results show that the complexity of the MSK-means algorithm is less than that of the K-means. This study also demonstrates that the MSK-means algorithm improves the classification accuracy by 4.7 % over the K-means, and has 0.7 % higher accuracy with 50 % less execution time than the SVM classifier using the half of the data as the training set. Hence, the MSK-means algorithm can be used efficiently for time series analysis and EEG classification.

In addition, the DPE index is applied to localize the epileptogenic zone from a public iEEG databases. The optimal delay factor of DPE are selected by analysing the different DPE indices between epileptic zone and non-epileptic zone from delay lag 0 to 50, then all DPE indices are forwarded to a K-means, MSK-means and SVM to identify the epileptogenic zone. The classification results show that the accuracy of detecting epileptogenic zone with the DPE index and MSK-means is higher than 93 % when the delay lag

(continued)

is 18. Hence, the DPE index can be a potentially suitable candidate for diagnostic protocol for epileptic EEG processing, especially with MSK-means classifiers. An implementation for R package of the DPE and MSK-means algorithms can be found in the MDPE R package (http://brain-graph.appspot.com/).

References

1. R.G. Andrzejak, K. Lehnertz, F. Mormann, C. Rieke, P. David, C.E. Elger, Indications of nonlinear deterministic and finite-dimensional structures in time series of brain electrical activity: dependence on recording region and brain state. Phys. Rev. E **64**(6), 061907 (2001)
2. R.G. Andrzejak, K. Schindler, C. Rummel, Nonrandomness, nonlinear dependence, and nonstationarity of electroencephalographic recordings from epilepsy patients. Phys. Rev. E **86**(4), 046206 (2012)
3. R.G. Andrzejak, K. Schindler, C. Rummel, The Bern-Barcelona EEG database, http://ntsa.upf.edu/downloads/andrzejak-rg-schindler-k-rummel-c-2012-nonrandomness-nonlinear-dependence-and. Accessed 10 Aug 2013
4. D. Arthur, S. Vassilvitskii, k-means++: the advantages of careful seeding, in *Proceedings of the Eighteenth Annual ACM-SIAM Symposium on Discrete Algorithms*, Philadelphia, 2007, pp. 1027–1035
5. B. Bahmani, B. Moseley, A. Vattani, R. Kumar, S. Vassilvitskii, Scalable k-means++. Proc. VLDB Endow. **5**(7), 622–633 (2012)
6. C. Bandt, B. Pompe, Permutation entropy: a natural complexity measure for time series. Phys. Rev. Lett. **88**(17), 174102 (2002)
7. E. Ben-Jacob, T. Doron, T. Gazit, E. Rephaeli, O. Sagher, V.L. Towle, Mapping and assessment of epileptogenic foci using frequency-entropy templates. Phys. Rev. E **76**(5), 051903 (2007)
8. J.H. Cho, H.C. Kang, Y.J. Jung, J.Y. Kim, H.D. Kim, D.S. Yoon, Y.H. Lee, C.H. Im, Localization of epileptogenic zones in Lennox-Gastaut syndrome using frequency domain source imaging of intracranial electroencephalography: a preliminary investigation. Physiol. Meas. **34**(2), 247–263 (2013)
9. M. Costa, A.L. Goldberger, C.K. Peng, Multiscale entropy analysis of complex physiologic time series. Phys. Rev. Lett. **89**(6), 068102 (2002)
10. G. Goffin, S. Dedeurwaerdere, K. Van Laere, W. Van Paesschen, Neuronuclear assessment of patients with epilepsy. Semin. Nucl. Med. **38**(4), 227–239 (2008)
11. L. Guo, D. Rivero, J. Dorado, C.R. Munteanu, A. Pazos, Automatic feature extraction using genetic programming: an application to epileptic EEG classification. Expert Syst. Appl. **38**(8), 10425–10436 (2011)
12. T.R. Henry, R.L. Van Heertum, Positron emission tomography and single photon emission computed tomography in epilepsy care. Semin. Nucl. Med. **33**(2), 88–104 (2003)
13. A. Karatzoglou, D. Meyer, K. Hornik, Support vector machines in R. J. Stat. Softw. **15**(9), 1–28 (2006)
14. R.C. Knowlton, R.A. Elgavish, A. Bartolucci, B. Ojha, N. Limdi, J. Blount, J.G. Burneo, L. Ver Hoef, L. Paige, E. Faught, P. Kankirawatana, K. Riley, R. Kuzniecky, Functional imaging: II. Prediction of epilepsy surgery outcome. Ann. Neurol. **64**(1), 35–41 (2008)
15. P. Krsek, M. Kudr, A. Jahodova, V. Komarek, B. Maton, S. Malone, I. Miller, P. Jayakar, T. Resnick, M. Duchowny, Localizing value of ictal SPECT is comparable to MRI and EEG in children with focal cortical dysplasia. Epilepsia **54**(2), 351–358 (2013)

16. J. Liu, G. Zhu, M. Xi, A k-means algorithm based on the radius. J. Guilin Univ. Electron. Technol. **33**(2), 134–138 (2013)
17. J.B. MacQueen, Some methods of classification and analysis of multivariate observations, in *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability*, Berkeley, 1967, pp. 281–297
18. M. Matilla-García, M. Ruiz Marín, Detection of non-linear structure in time series. Econ. Lett. **105**(1), 1–6 (2009)
19. F. Mormann, R.G. Andrzejak, C.E. Elger, K. Lehnertz, Seizure prediction: the long and winding road. Brain **130**(2), 314–333 (2007)
20. N. Nicolaou, J. Georgiou, Detection of epileptic electroencephalogram based on permutation entropy and support vector machines. Expert Syst. Appl. **39**(1), 202–209 (2012)
21. U. Orhan, M. Hekim, M. Ozer, EEG signals classification using the K-means clustering and a multilayer perceptron neural network model. Expert Syst. Appl. **38**(10), 13475–13481 (2011)
22. K. Polat, S. Günes, Classification of epileptiform EEG using a hybrid system based on decision tree classifier and fast Fourier transform. Appl. Math. Comput. **187**(2), 1017–1026 (2007)
23. A. Popov, O. Avilov, O. Kanaykin, Saturation of electroencephalogram permutation entropy for large time lags, in *IEEE XXXIII International Scientific Conference Electronics and Nanotechnology (ELNANO)*, Kiev, 16–19 Apr 2013, pp. 251–254
24. F. Rosenow, H. Lüders, Presurgical evaluation of epilepsy. Brain **124**(9), 1683–1700 (2001)
25. S. Siuly, Y. Li, P. Wen, Clustering technique-based least square support vector machine for EEG signal classification. Comput. Methods Programs Biomed. **104**(3), 358–372 (2011)
26. Y. Song, P. Liò, A new approach for epileptic seizure detection: sample entropy based feature extraction and extreme learning machine. J. Biomed. Sci. Eng. **3**(6), 556–567 (2010)
27. M.V. Spanaki, S.S. Spencer, M. Corsi, J. MacMullan, J. Seibyl, I.G. Zubal, Sensitivity and specificity of quantitative difference SPECT analysis in seizure localization. J. Nucl. Med.: Off. Publ. Soc. Nucl. Med., **40**(5), 730–736 (1999)
28. A. Subasi, Automatic detection of epileptic seizure using dynamic fuzzy neural networks. Expert Syst. Appl. **31**(2), 320–328 (2006)
29. P. van Mierlo, E. Carrette, H. Hallez, R. Raedt, A. Meurs, S. Vandenberghe, D. Van Roost, P. Boon, S. Staelens, K. Vonck, Ictal-onset localization through connectivity analysis of intracranial EEG signals in patients with refractory epilepsy. Epilepsia **54**(8), 1409–1418 (2007)
30. A. Vattani, k-means requires exponentially many iterations even in the plane. Discret. Comput. Geom. **45**(4), 596–616 (2011)
31. J. Wu, W. Sutherling, S. Koh, N. Salamon, R. Jonas, S. Yudovin, R. Sankar, W. Shields, G. Mathern, Magnetic source imaging localizes epileptogenic zone in children with tuberous sclerosis complex. Neurology **66**(8), 1270–1272 (2006)
32. G. Zhu, Y. Li, P.P. Wen, Analysing epileptic EEGs with a visibility graph algorithm, in *5th International Conference on Biomedical Engineering and Informatics (BMEI2012)*, Chongqing, 16–18 Oct 2012, pp. 432–436
33. G. Zhu, Y. Li, P.P. Wen, S. Wang, M. Xi, Epileptogenic focus detection in intracranial EEG based on delay permutation entropy. AIP Conf. Proc. **1559**, 31–36 (2013)
34. G. Zhu, Y. Li, P. Wen, S. Wang, N. Zhong, Unsupervised classification of epileptic EEG signals with multi scale K-means algorithm, in *Brain and Health Informatics*, Maebashi, ed. by K. Imamura, S. Usui, T. Shirao, T. Kasamatsu, L. Schwabe, N. Zhong. Lecture Notes in Computer Science, vol. 8211 (Springer, Cham, 2013), pp. 158–167

Chapter 9
Tracking of EEG Activity Using Motion Estimation to Understand Brain Wiring

Humaira Nisar, Aamir Saeed Malik, Rafi Ullah, Seong-O Shim, Abdullah Bawakid, Muhammad Burhan Khan, and Ahmad Rauf Subhani

Abstract The fundamental step in brain research deals with recording electroencephalogram (EEG) signals and then investigating the recorded signals quantitatively. Topographic EEG (visual spatial representation of EEG signal) is commonly referred to as brain topomaps or brain EEG maps. In this chapter, full search block motion estimation algorithm has been employed to track the brain activity in brain topomaps to understand the mechanism of brain wiring. The behavior of EEG topomaps is examined throughout a particular brain activation with respect to time. Motion vectors are used to track the brain activation over the scalp during the activation period. Using motion estimation it is possible to track the path from the starting point of activation to the final point of activation. Thus it is possible to track the path of a signal across various lobes.

Keywords Brain activation • EEG • Topomaps • Motion estimation • Full search

H. Nisar (✉) • M.B. Khan
Faculty of Engineering and Green Technology, Department of Electronic Engineering, Universiti Tunku Abdul Rahman, Kampar, Perak, Malaysia
e-mail: humaira@utar.edu.my

A.S. Malik • A.R. Subhani
Department of Electrical and Electronic Engineering, Centre for Intelligent Signal and Imaging Research, Universiti Teknologi PETRONAS, Tronoh, Perak, Malaysia
e-mail: aamir_saeed@petronas.com.my

R. Ullah
Comsats Institute of Information Technology, Islamabad, Pakistan

S.-O. Shim • A. Bawakid
Faculty of Computing and Information Technology, King Abdul Aziz University, North Branch, Jeddah, Kingdom of Saudi Arabia

C. Sun et al. (eds.), *Signal and Image Analysis for Biomedical and Life Sciences*, Advances in Experimental Medicine and Biology 823,
DOI 10.1007/978-3-319-10984-8_9

9.1 Introduction

The motion information that is being extracted from a sequence of 2D images has a number of applications in the field of image processing, medical image investigation/analysis, object tracking, remote sensing, and video compression. Estimating the motion present in a video sequence using the motion vectors (MV) is called motion estimation. Hence using motion estimation it is possible to track the motion of an individual object or a group of objects in a video sequence [4, 9, 12]. Electroencephalography (EEG) is the recording of electrical activity along the scalp. The flow of current due to firing of neurons in the brain results in the voltage fluctuation that is measured as EEG. The visual image of brain changes with the change in the activation of brain. It means that images of brain taken at regular intervals will be different. Hence if consecutive brain images are acquired then we can observe the changes in the images. The changes in the images may correspond to some motion pattern that may be tracked or estimated. Hence, motion estimation techniques can be used to detect the changes in activation. The spatio-temporal correlation between consecutive frames in the sequence can be exploited to find the direction of motion and hence the flow of signal across various lobes in the human brain.

In this chapter, we are focusing on the motion vectors that are created from the EEG signal movement due to brain activity in the topomap sequence, and exploit these motion vectors for further analysis. Our video sequences will consist of topomaps generated from the EEG data in our experiments. Our key contribution is to exploit motion estimation algorithms for brain topomap analysis so that we can understand the mechanism of signal flow in the brain under certain activity. We will use full search (FS) block matching algorithm (BMA) for estimating the motion as this algorithm gives good estimation. Optical flow techniques are also used for motion estimation. However, we did not consider them due to their high computational complexity because they cannot be used for many real time applications of EEG. Although full search motion estimation algorithm has high computational complexity too, there are a large variety of BMA methods with very low computational complexity that are suitable for real time applications. The aim of this chapter is to provide the proof of concept for tracking EEG activity using FS BMA method. In future, we plan to utilize and optimize fast BMA techniques for real time processing.

Thus our goal in this chapter is three-fold: First to analyze the behavior of different brain lobes throughout a particular brain activity with respect to time. This can be done by tracking the path of motion across the brain lobes using motion vectors. Secondly, to track the paths followed by EEG signals during the overall activity and thirdly, to select the optimal path followed. For this purpose, we employ the full search BMA, which is the best among all other BMAs with respect to accuracy.

We organize the rest of the chapter as follows. In Sect. 9.2, we will explain the EEG signals and brain topomaps respectively. In Sect. 9.3, we will discuss motion

estimation and compare some fast motion estimation techniques. In Sect. 9.4, we will analyze topomaps using full search algorithm. Results and discussions will be presented in Sect. 9.5. This is followed by conclusion and references.

9.2 EEG Signals

EEG signals are obtained by recording the electrical activity of the brain. EEG is recorded by EEG machines, which include electrodes, amplifiers, filters, recording units, and displaying devices. The electrical activity recorded is the reflection of activity from the brain structure underneath the cortex [19]. In 1929; Hans Berger made the first recording from the human scalp, while in 1870, similar studies were carried out on animals. The nerve cells in the brain continuously create small electrical signals called action potentials. The nerve cells transmit the information throughout the body electrically. With the help of neurotransmitter, the action potentials move from one cell to another across a synaptic gap. The electrodes pick up the small electrical brainwaves produced by the cells and pass through the amplifier that amplifies them enough to be displayed. Different EEG machines may have different number of channels, e.g., 8, 16, 24, 64, 128 etc. Different mind states have different EEG activations e.g., sleep, alertness, dreaming etc. which are associated with different brain waves, such as, theta, alpha, delta, beta, and gamma. EEG activity is measured in micro-volts and main frequencies of interest are up to 100 Hz. Different brain rhythms have different frequencies and amplitude [1], that are explained as follows:

Delta waves (<4 Hz): These waves have the highest amplitudes and lowest frequencies. These are primarily associated with deep sleep.

Theta waves (4–8 Hz): These waves are observed in the adults and older children when there are memory intensive activities. This pattern also appears in deep meditation and light sleep. It also includes rapid eye movement (REM) sleep state.

Alpha waves (8–14 Hz): Alpha waves are associated with wakefulness, closing the eye, effortless alertness, and creativity.

Beta waves (14–30 Hz): These waves are found only in normal adults. These waves correlate with active attention, active thinking, solving critical problems or focus on the outside world. Therefore, these are also known as a Sensory Motor Rhythm.

Gamma waves (>30 Hz): These waves are associated with high-level information processing. Usually, these are not of medical interest and often filtered out from EEG signals.

Various applications of EEG have been identified in the existing brain research like biomedical data analysis, brain computer interface, and military applications [3, 7, 16]. The most common application of EEG in medicine is to identify the diseases like seizure disorder, head injuries, and brain tumors etc. EEG analysis also

helps us in solving problems like sleep disorder, changes in behavior, psychological and stress related diseases etc. EEG can also be used for quantitative measurement of brain activities before liver or heart transplantation.

9.3 EEG Topomaps

Brain mapping is the visual-spatial illustration of biological properties of brain known as topographic maps or topomaps [10,17]. The parameter under study is plotted onto a picture of the brain. Amplitude at a given frame is usually represented as strength/intensity, and amplitudes at undetermined frames are estimated to present a smooth display. Features of EEG are highlighted by using these displays, which are referred as EEG topomaps. The brain topomaps should not be confused with cortical mapping of brain that is acquired by direct electrical cortical stimulation or with the brain mapping that is acquired by applying neuro imaging techniques. There is no direct relationship of EEG with these techniques. Brain is divided into different lobes based on its functions as is given in Fig. 9.1. One of the twenty-nine topomaps of the brain that have been used in our experiments is shown in Fig. 9.1. The four lobes are represented by different colors for the purpose of illustration only. The four lobes are frontal, parietal, temporal (left temporal region (LTR), right temporal region (RTR)), and occipital. The deep fold between frontal and parietal lobes is referred to as central motor region (CMR). The location and function of various lobes are as follows:

- **Frontal lobe** is ahead of the central sulcus and is responsible for planning, thinking, and emotions.
- **Parietal lobe** is behind the central sulcus and is involved in sensation of temperature, taste, touch, pain, pressure, etc.
- **Temporal lobe** is below the lateral fissure, and is responsible for the auditory sense. Specific tones and the sound strength is recognized by temporal lobe. Auditory damage and deafness can occur due to tumor or accidental injuries in this area. In addition, this area plays a vital role in emotion processing and memory.
- **Occipital lobe** is located behind the parietal and temporal lobes and deals with visual information processing. Because of its location, this lobe is less susceptible

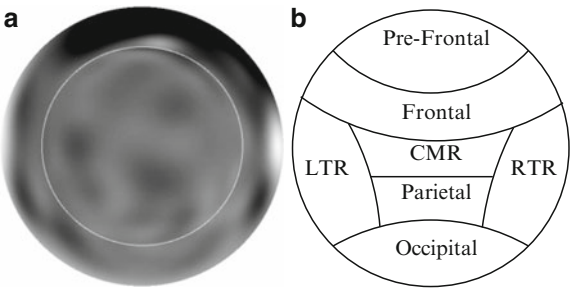


Fig. 9.1 (**a**) Brain topomap. (**b**) The corresponding map with different brain lobes

to injuries, however significant trauma in this area may result in visual perceptual system disorder or even complete blindness.

- **Central lobe/CMR** is in the central area, i.e., located at the intersection of the cerebral hemisphere with the brain stem. CMR separates the frontal lobe from the parietal lobe. This is responsible for emotional and sexual aspects.

9.4 Motion Estimation

Motion estimation refers to the estimation of the displacement from one frame to another frame in the image sequences. Consecutive video frames are spatially and temporally correlated with each other. The estimation of displacement is done by obtaining the motion vector. These estimated motion vectors eventually represent the estimated motion between the consecutive frames. Hence the temporal redundancy between the consecutive frames is eliminated by employing motion estimation and compensation. The spatio-temporal correlation and the directions of motion vectors are used to estimate the motions in the video [13–15]. In a normal video, motion can be a complex combination of rotation and translation. It is quite difficult to estimate these motions and may require large amount of processing [2]. However, translational motion is easily estimated and is used successfully for motion compensation. In this research unlike the traditional video, we use the topomaps of EEG to represent video sequences, which is a complex combination of scaling, i.e., expansion and shrinking, along with slight translation, rotation and shifting. The motion vectors in case of EEG topomaps will represent expansion and contraction of the regions instead of object movements in the normal video.

The two commonly used techniques for motion estimation are pel-recursive algorithms (PRA) [11] and block matching algorithms (BMA) [5]. PRAs are based on individual pixels while BMAs are based on rectangular blocks where all the pixels within a block are assumed to have the same motion activity. PRAs are computationally complex while BMAs are fast and suitable for simple realization.

In block matching based motion estimation (BBME), the current frame is partitioned into square blocks of pixels and the best match of these blocks is found inside the reference frame using a predefined distortion criterion. The best match is then used as a predictor for the block in the current frame, whereas the displacement between the two blocks is usually defined as the motion vector (MV), which is associated with the current block [14]. Full search (FS) is the most straightforward and optimal BMA which searches exhaustively inside the search window to find the MV. Despite very heavy computations required in FS it is widely used in video coding applications due to its simplicity. Several fast block matching motion estimation algorithms have been proposed so far. These fast algorithms involve approaches like unimodal error surface assumption (UESA), variable search range instead of fixed one, methods using multi-resolution, spatial and temporal correlation of MVs, pixel decimation etc.

9.4.1 Full-Search (FS) Block Matching Algorithm

This BMA is very simple but it involves high computational cost, where the cost function is calculated at each possible location in a search window. There are many fast BMAs but we have to make a trade-off between efficiency and image quality. In this research, our main focus is to obtain accurate motion vectors. Hence the computational cost of FS is acceptable. The block within the search window used in FS is depicted in Fig. 9.2 and the block diagram for FS is given in Fig. 9.3. The current frame is divided into blocks of size $N \times N$, and for every block in the current frame, the previous (reference) frame is searched within the search space to find out the closest match as shown in Fig. 9.3. The parameter p is the search range. Figure 9.4 shows the blocks and the corresponding search area for $N = 3$ and $p = 2$. In Fig. 9.4, $m(i, j)$ and $s(i, j)$ represent each pixel in the block and search area respectively. The search area consists of $(N + 2p) \times (N + 2p)$ pixels.

In order to measure the similarity between current frame block and a candidate block of the reference frame, various distortion measure criteria may be used, such as minimum mean square error (MSE), minimum mean absolute difference (MAD), or minimum sum of absolute differences (SAD). Some of the distortion measures are defined as follows:

$$MSE = \frac{1}{N^2} \sum_{i=0}^{N-1} \sum_{j=0}^{N-1} (C_{ij} - R_{ij})^2 \tag{9.1}$$

$$MAD = \frac{1}{N^2} \sum_{i=0}^{N-1} \sum_{j=0}^{N-1} |C_{ij} - R_{ij}| \tag{9.2}$$

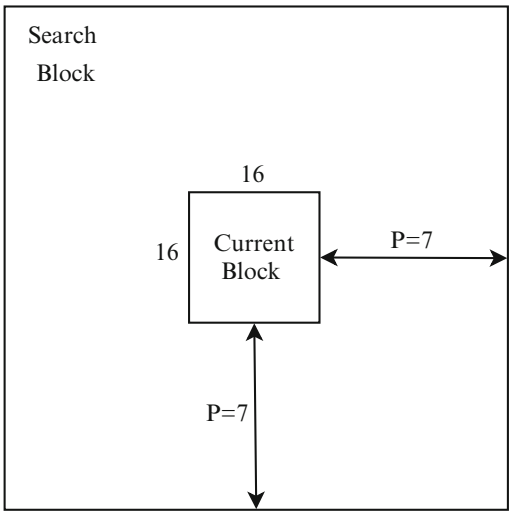


Fig. 9.2 Search window for FS block motion estimation algorithm

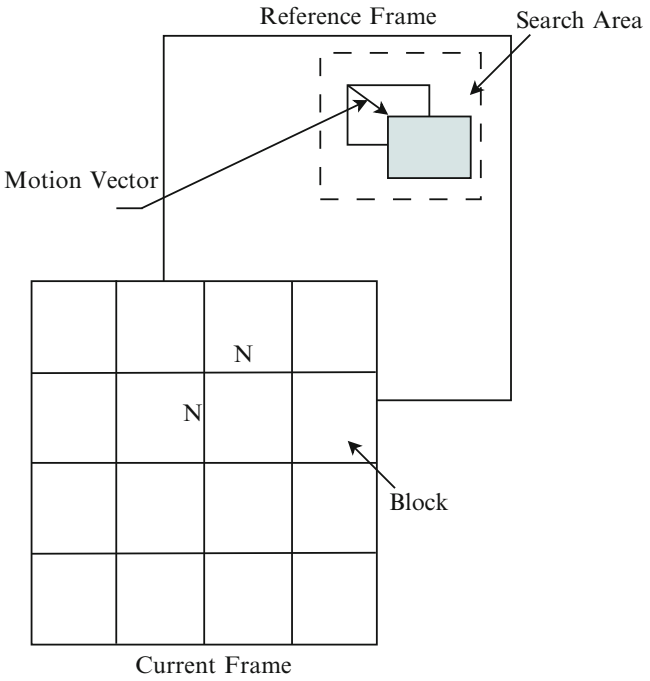


Fig. 9.3 Block diagram for a full search algorithm

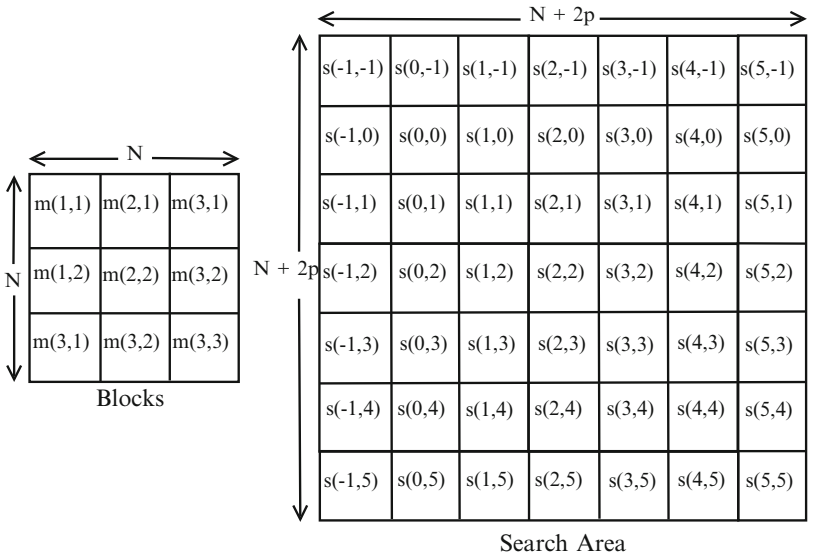


Fig. 9.4 A sample blocks and search area with $N = 3$ and $p = 2$

where N is the size of block, C_{ij} and R_{ij} are the pixels that are compared in the current and reference blocks respectively.

Structural similarity index measure (SSIM) and peak signal to noise ratio (PSNR) are used to evaluate the performance of the BMAs. The higher PSNR means the better match and vice versa. FS gives highest PSNRs as compared to all other BMAs. PSNR and SSIM are defined as follows:

$$PSNR = 10 \log_{10} \frac{255^2}{MSE} \tag{9.3}$$

$$SSIM(x, y) = \frac{(2\mu_x\mu_y + c_1)\,(2\sigma_{xy} + c_2)}{(\mu_x^2 + \mu_y^2 + c_1)\,(\sigma_x^2 + \sigma_y^2 + c_2)} \tag{9.4}$$

where x and y are two windows of size $N \times N$, and μ_x and μ_y are the average of x and y respectively. σ_x^2 and σ_y^2 are the variances of x and y respectively. $c_2 = (k_2 L)^2$ with L being the maximum value, i.e., $2^{\#bits}$, and $k_1 = 0.01$ and $k_2 = 0.03$ by default.

There is no precise rule to select either PSNR or SSIM measurement for evaluating the image/frames quality. PSNR is the peak signal to noise ratio whereas SSIM shows the similarity measurements between two frames. The PSNR between the compensated and the corresponding reference frame has been calculated. The compensated frames are generated by using motion vectors of original and the reference frames.

The followings are a brief overview of fast motion estimation algorithms.

9.4.2 Three Step Search (TSS) Algorithm

TSS was introduced by Koga et al. in 1981, and become very popular for its simplicity, robustness and near to optimal solution [6]. This algorithm is one of the earliest BMAs. In this BMA, an initial step size is selected and eight points at a distance of selected step size around the search centre are selected for comparison. In the next step, the step size is bisected and the centre is moved to the point with the minimum cost/distortion. This procedure is repeated until the step size becomes equal to one. The main problem with TSS is the use of uniformly allocated checking point pattern in the first step, where eight points around the center point at the distance of fixed step size are chosen for comparison. Hence TSS is inefficient for small motion estimation.

9.4.3 New Three Step Search (NTSS) Algorithm

In TSS, the step size in the first step is large and is not capable of detecting small motion. This problem is solved in NTSS that employs the center-biased checking point pattern in the first step. NTSS has better results than TSS and the computational complexity is reduced by adopting a half-stop search technique [8].

9.4.4 Four Step Search (FSS) Algorithm

FSS adopts a center biased search pattern with a nine point comparison in the first step. In the next steps either three or five points are selected based on the minimum distortion in the previous step. The search is stopped if the search center is chosen as the minimum distortion point and step size is reduced to 1. In the last step eight points are searched [18].

9.4.5 Diamond Search (DS) Algorithm

This algorithm is very similar to FSS, but the search point pattern is changed to diamond from square. The number of search steps is unlimited. Two different patterns are used in DS: large diamond search pattern (LDSP) and small diamond search pattern (SDSP) [20]. DS algorithm outperforms the TSS and FSS algorithm. It produces very similar results to NTSS but the computational cost is reduced in DS.

9.4.6 Two Dimensional Logarithmic Search (TDLS) Algorithm

TDLS is very similar to TSS. It requires more steps than TSS but it can be more accurate, especially in the case of large search window [5]. TDLS is explained as follows: Pick an initial step size and look at the central point and the four points at some pre-defined distance from central point on both axes. If the best match is at the centre then decrease the step size to half otherwise the best match will become the centre and the same procedure is repeated. When the step size becomes one, eight neighboring points are chosen for the search and the best among them is picked.

9.4.7 Orthogonal Search Algorithm (OSAlg)

OSAlg is an amalgam of TSS and TDLS. The optimal point is found by searching the vertical pattern and horizontal pattern respectively [5]. The procedure of OSAlg is described as follows: Two points at a pre-defined distance from the search center are picked in the horizontal direction and a point with minimum distortion is chosen. The center is moved to this point. Then two points at the same step size from the centre are chosen in the vertical direction and the point with the minimum distortion is chosen. The procedure is continued till the step size is one.

Figure 9.5 shows a comparison of FS and several fast ME algorithms in terms of PSNR. From the comparison it is clear that FS performs better than all fast algorithms in terms of the picture quality. Hence in the rest of the chapter we will only discuss results with FS algorithm for the brain topomaps analysis.

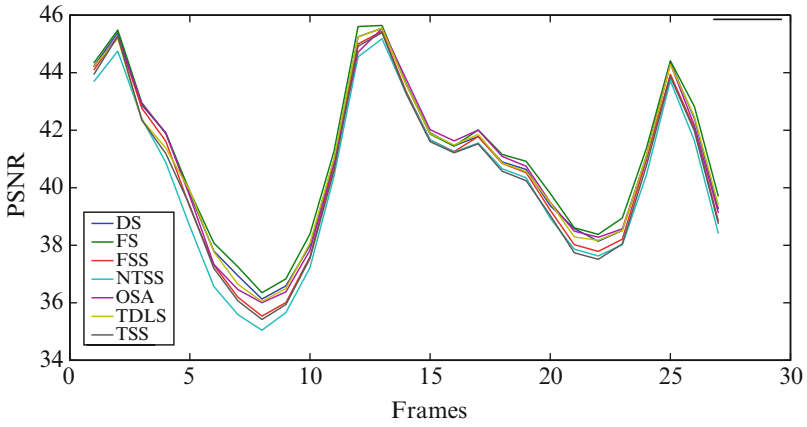


Fig. 9.5 PSNR performance of several block matching algorithms

9.4.8 Tracking Brain Activity Using FS Motion Estimation Algorithm

In this chapter, we are using FS block matching algorithm to track the pattern of activity in the EEG topomaps and investigate these patterns. By applying the FS algorithm, the motion vectors are identified in their respective search spaces. Assume that during a certain activity, the activation starts from the occipital lobe (originating area) and goes towards the frontal lobe (target area) in a particular direction. Initially there may be several movements in several directions from the starting area to the target area. At every frame, we check the status of these directions by the help of motion vectors. In most cases the motion vectors cross the boundaries of the lobes while moving towards the target area. In this way, we can track the EEG signals throughout the activity. Secondly, we track the optimum route from the starting area to the target area. The optimum route is estimated based on number of frames it takes to reach the final destination. If the motion is trapped in an area, then those motion vectors are not considered for calculating the path followed. The length of the path, from starting point to final point, is based on those motion vectors that point away from the local area and are called candidate motion vectors (CMVs). Starting from the CMV all the vectors are counted in calculating the length of the path. It means that we may ignore only those vectors which are trapped in the starting area; otherwise, they are considered.

9.5 Experimental Setup and Data Acquisition

The EEG topomaps have been collected using a 128 channel EGI (Electrical Geodesics, Inc.) EEG machine. The resolution is 24 bits per sample at a sampling rate of 250 samples per second. This data was collected from one subject while

performing two different experiments. In the first case a simple mathematical problem is solved, by seeing the problem on the computer screen. The area of the brain that is used for this problem is occipital where the subject process what he sees on the screen. The language areas (the very left temporal, surrounding the left ear) are also used while performing the mathematical calculation.

For the second experiment the EEG data was recorded while the subject was pressing the ENTER button of the keyboard by seeing a blue box on the screen. Thus, the activation is expected to start from the occipital lobe. The recorded data was then converted into the topographic format, i.e., EEG topomaps.

The experiments are performed on 29 topomaps (frames). The sequence of these frames collectively represents a 5 s video, where each frame takes 172.4 ms. The frame difference between the reference and current frame is set to 2. The frame size is 256×256 pixels, the block size is set to 8×8 and the search window size p is set to 7 in our experiments for the FS algorithm. The algorithms have been simulated using MATLAB programming language.

9.6 Discussion of Results

One of the topomaps (frames) and its corresponding predicted frame is shown in Fig. 9.6. The motion vectors show the direction of EEG signals between the two frames, i.e., at a time interval t and time interval $t + 1$. It can be observed that there are several movements within a frame interval. The motion vectors show the direction of brain activation from one location to another.

In the ordinary video, the objects are usually moving, and the major parts of the frame are not moving. However, in EEG topomaps, most parts of the frame are moving and these movements are quite complex as compared to the movements of objects in common video. In topomaps, most of the areas are expanding and shrinking. Figure 9.6c shows many vectors, where some of the vectors indicate the shrinking and some of them indicate the expansions within the frame. Figure 9.7 shows the brain topomaps with and without background for frame 2, 12, and 22 out of 29 frames. The lower row clearly indicates the motion vector positions/directions with respect to the lobes. These vectors show the movement of EEG activity through different lobes.

Figure 9.8 shows the movement of the motion vectors, where each vector is tracked from the starting point to indicate the path followed throughout the activity. Each vector is under consideration individually with respect to its location. It is observed from Fig. 9.8 that in some areas, the movement is trapped within the local region for partial duration, or entire duration of activation. This trapping can be either in the initial, final, or in the middle stage of activation. We are taking into account all the paths from the starting points to the target points. Figure 9.8a–h show different paths followed by the respective motion vectors. At this stage, we are considering every type of movement either local or not. Later on, we will exclude the movements within the local area. These local movements are ignored, because we are interested in those movements that result in the crossing of signal to another lobe.

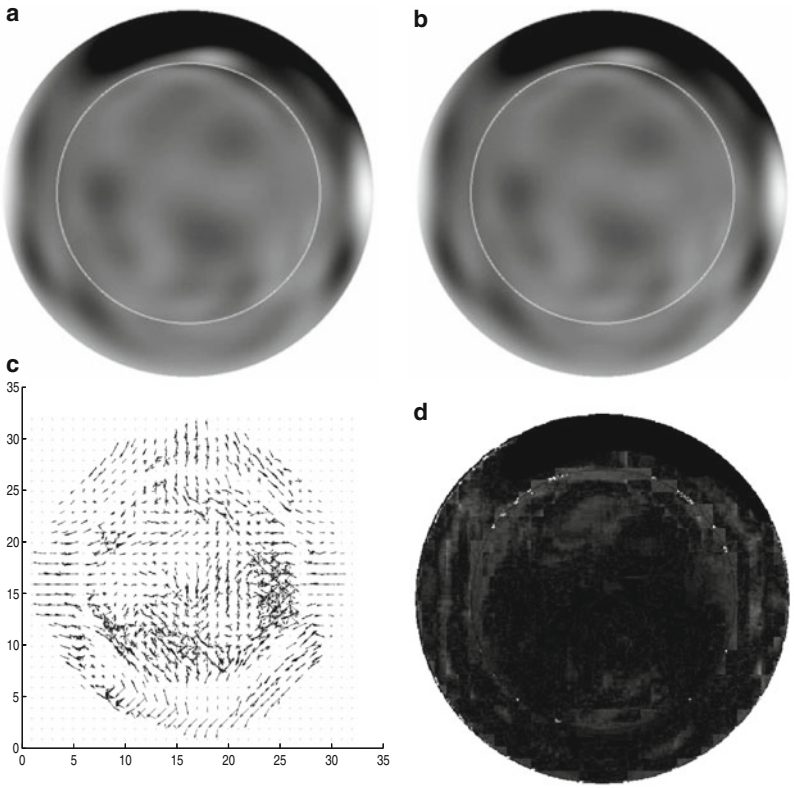


Fig. 9.6 Sample topomaps (**a**) reference frame (**b**) predicted frame (**c**) motion vectors (**d**) associated prediction error

Figure 9.8d, f shows that the motion is trapped on the border of the topomap. Thus, it is very important to discriminate such type of vectors from all other vectors. The vectors starting from a single location (individual path) have been taken under consideration in each frame instead of showing all the vectors as shown in Fig. 9.9.

Movements within the local area are ignored and inter lobe movements are considered for analyzing the vectors with respect to time. The path with lesser number of candidate motion vectors (CMVs) is selected as a optimal path. The optimal routes from the starting point to the ending points are based on counting the time intervals while ignoring the vectors lying in the local area.

Conclusions

In this chapter, we have used full search (FS) block motion estimation (BME) algorithm to analyze and track the motion in the EEG topomaps that result from brain activation based on the stimuli. Using the motion vectors, we

(continued)

examine the behavior of the brain and can specify the areas of the brain that are activated because of the stimulus. Further, we can track the path followed by individual vectors that represent the brain activation path. Finally we have tracked the optimal path followed during an activity. Hence using FS algorithm we are able to illustrate the movements in the brain under certain activity and can display these movements using motion vectors. In addition we can also follow the path throughout the activity. Finally we may be able to select the optimal path for a certain activity. Hence this research can help to understand the mechanism of the information flow in the brain which can lead to the development of brain computer interface systems.

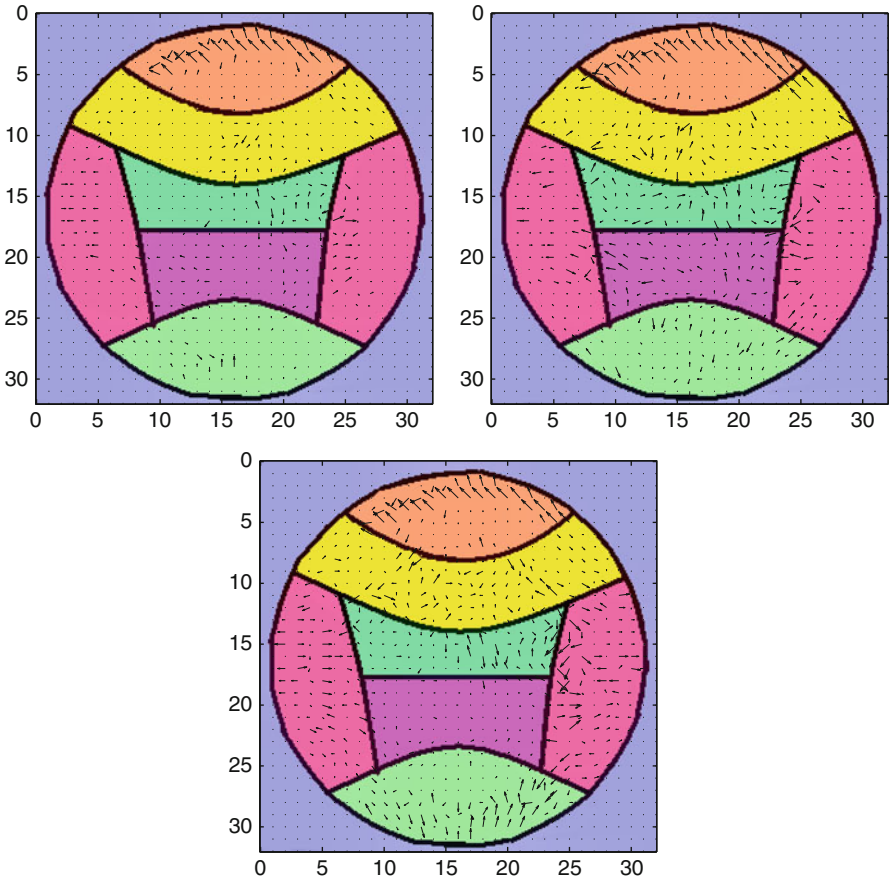


Fig. 9.7 Motion vectors showing EEG activation direction for frame no 2, 12, and 22

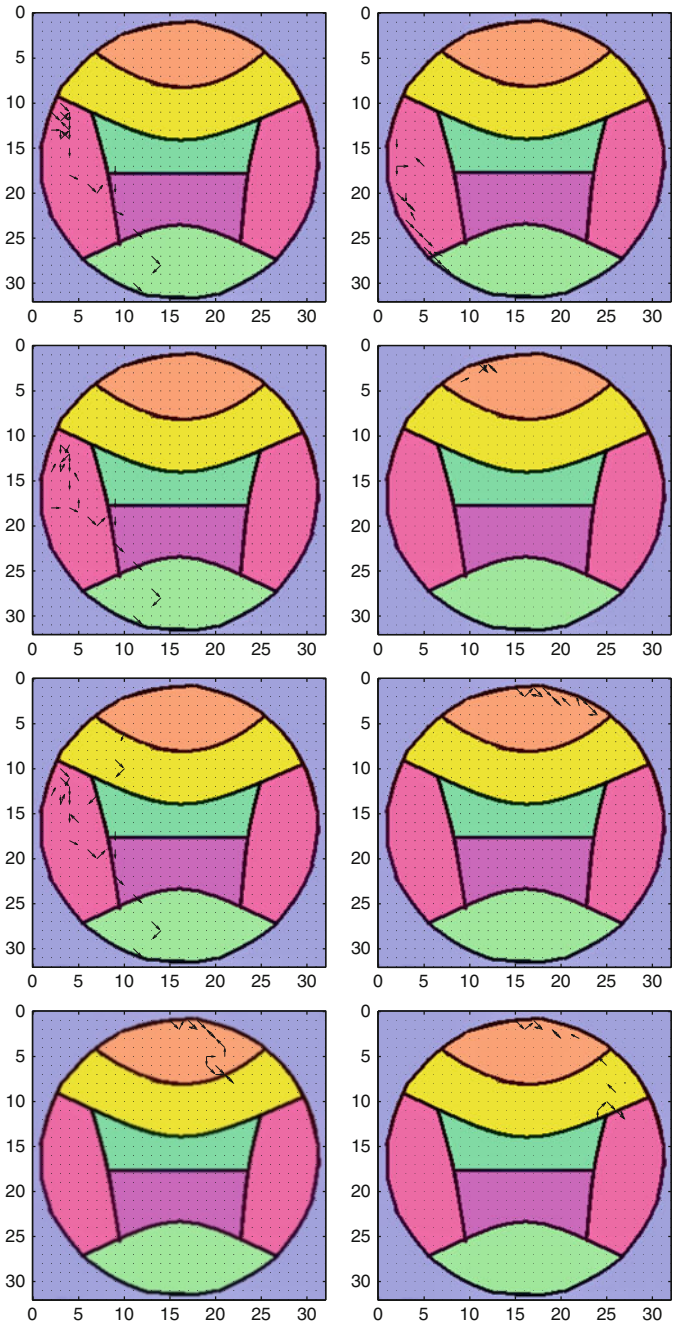


Fig. 9.8 The directions of motion vectors and their respective locations while pressing the ENTER button of the keyboard

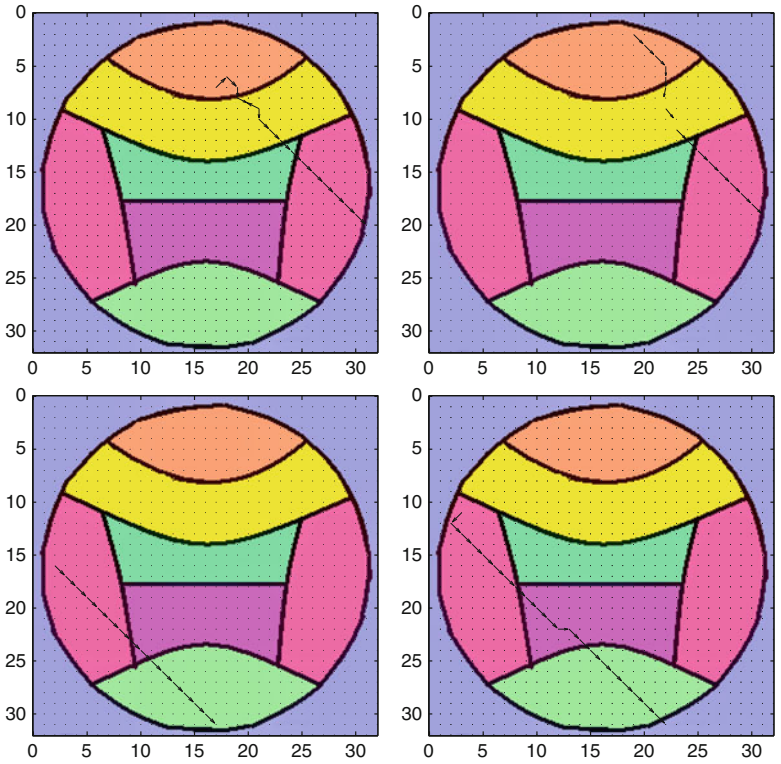


Fig. 9.9 The direction of motion vectors and their respective location during simple MATH problem solving

Acknowledgements This work was funded by the Deanship of Scientific Research (DSR), King Abdulaziz University, Jeddah, under grant No. 968-009-D1434. The authors, therefore, acknowledge with thanks DSR technical and financial support.

References

1. V.P.C. Babiloni, C.D. Gratta, A. Ferretti, G.L. Romani, Fundamentals of electroencefalography, magnetoencefalography, and functional magnetic resonance imaging. Int. Rev. Neurobiol. **86**, 67–80 (2009)
2. K.S. Bhushan, Motion based foreground segmentation using block matching algorithm. Int. J. Recent Trends Eng. **1**, 158–160 (2009)
3. T.F. Collura, Applications of small brainwave machines, in *Proceedings of Annual Conference on Brain Modification and EEG*, Palm Springs, Feb 1997
4. M. Djalalov, H. Nisar, Y. Salih, A.S. Malik, An algorithm for vehicle tracking and detection, in *Proceedings of International Conference on Intelligent and Advanced Systems*, Kuala Lumpur, 2010

5. J. Jain, Displacement measurement and its application in interframe image coding. IEEE Trans. Commun. **29**, 1799–1808 (1981)
6. T. Koga, K. Iinuma, A. Hirano, Y. Iijima, T. Ishiguro, Motion compensated interframe coding for video conferencing, in *Proceedings of NTC81*, New Orleans, 1981
7. W.T. Lee, H. Nisar, A.S. Malik, K.H. Yeap, A brain computer interface for smart home control, in *Proceedings of IEEE International Symposium on Consumer Electronics*, Hsinchu, 3–6 June 2013, pp. 35–36
8. R. Li, B. Zeng, M.L. Liou, A new three-step search algorithm for block motion estimation. IEEE Trans. Circuits Syst. Video Technol. **4**, 438–442 (1994)
9. X. Marichal, Motion estimation and compensation for very low bitrate video coding, PhD thesis, Laboratoire de Telecommunication et Teledetection, Université Catholique de Louvain, 1998
10. M.R. Nuwer, Quantitative EEG: I. Techniques and problems of frequency analysis and topographic mapping. J. Clin. Neurophysiol. **5**(1), 1–44, (1988)
11. A. Netravali, J. Robbins Motion-compensated television coding: part I. Bell Syst. Tech. J. **58**, 631–670 (1979)
12. H. Nisar, T.S. Choi, An advanced center biased search algorithm for motion estimation, in *International Conference on Image Processing*, Vancouver, vol. 1, 2000, pp. 832–835
13. H. Nisar, T.S. Choi, Fast motion estimation algorithm based on spatio-temporal correlation and direction of motion vectors. Electron. Lett. **42**, 1384–1385 (2006)
14. H. Nisar, T.S. Choi, Multiple initial point prediction based search pattern selection for fast motion estimation. Pattern Recognit. **42**, 475–486 (2009)
15. H. Nisar, A.S. Malik, T.S. Choi, Content adaptive fast motion estimation based on spatio-temporal homogeneity analysis and motion classification. Pattern Recognit. Lett. **33**, 52–61 (2012)
16. H. Nisar, H.C. Balasubramaniam, A.S. Malik, Brain computer interface for operating a robot, in *Proceedings of International Symposium on Computational Models for Life Sciences*, Sydney. AIP conference proceedings, vol. 1559 (2013)
17. C.M. Pechura, J.B. Martin, *Mapping the Brain and Its Functions: Integrating Enabling Technologies into Neuroscience Research*. (The National Academies Press, Washington, D.C., 1991)
18. L.M. Po, W.C. Ma, A novel four-step search algorithm for fast block motion estimation. IEEE Trans. Circuits Syst. Video Technol. **6**, 313–317 (1996)
19. E.J. Smith, Introduction to EEG, http://www.ebme.co.uk/articles/clinical-engineering/56-introduction-to-eeg. Jan 2014
20. S. Zhu, K.K. Ma, A new diamond search algorithm for fast block-matching motion estimation. IEEE Trans. Image Process. **9**, 287–290 (2000)

Part II
Image Analysis

Chapter 10
Towards Automated Quantitative Vasculature Understanding via Ultra High-Resolution Imagery

Rongxin Li, Dadong Wang, Changming Sun, Ryan Lagerstrom, Hai Tan, You He, and Tiqiao Xiao

Abstract This chapter presents an approach to processing ultra high-resolution, large-size biomedical imaging data for the purposes of detecting and quantifying vasculature and microvasculature. Capturing early signs of any changes in vasculature may have significant values for early-diagnosis and treatment assessment due to the well understood observation that vascular changes precede cancerous growth and metastasis. With the advent of key enabling technologies for extremely high-resolution imaging, such as synchrotron radiation based computed tomography (CT), the required levels of detail have become accessible. However, these technologies also present challenges in data analysis. This chapter aims to offer some insights as to how these changes might be best dealt with. We argue that the necessary steps in quantitative understanding of vasculatures include targeted data enhancement, information reduction aimed at characterizing the linear structure of vessels, and quantitatively describing the vessel hierarchy. We present results on cerebral and liver vasculatures of a mouse captured at the Shanghai Synchrotron Radiation Facility (SSRF). These results were achieved with a processing pipeline comprising of our empirically selected component for each of the above steps. Towards the end, we discuss how alternative and additional components may be incorporated for improved speed and robustness.

Keywords Angiogenesis and neovascularization • Vasculature quantification • Vessel skeletonization • Synchrotron imagery • Vascular hierarchy

R. Li (✉) • D. Wang • C. Sun • R. Lagerstrom
Digital Productivity Flagship, CSIRO, Locked Bag 17, North Ryde, Sydney, NSW 1670, Australia
e-mail: ron.li@csiro.au; dadong.wang@csiro.au; changming.sun@csiro.au; ryan.lagerstrom@csiro.au

H. Tan • Y. He • T. Xiao
Shanghai Synchrotron Radiation Facility (SSRF), Chinese Academy of Sciences, Shanghai Institute of Applied Physics, 239 Zhangheng Road, Pudong District, Shanghai 201204, China
e-mail: tanhai@sinap.ac.cn; heyou@sinap.ac.cn; tqxiao@sinap.ac.cn

C. Sun et al. (eds.), *Signal and Image Analysis for Biomedical and Life Sciences*, Advances in Experimental Medicine and Biology 823,
DOI 10.1007/978-3-319-10984-8_10

10.1 Introduction

Detection of malignant tumor growth at an early stage is crucial to satisfactory prognosis and patients' survival. Equally important is the assessment of the efficacy of various treatment modalities, including pharmaceutical (particularly angiogenesis inhibitors or anti-angiogenic agents), surgical, chemotherapeutical and radiation interventions. For both these purposes, the accurate assessment of vascular and microvasculature changes (such as microvasculature angiogenesis) can play a critical and indispensable role. This is because, as is well understood, tumor growth and metastasis require markedly increased amounts of nutrients, and are thus preceded and accompanied by angiogenesis and neovascularization [17]. It is plausible to stipulate that any changes in vasculature are likely to be minute in the early stages, and require a high resolution image for the detection and measurement of any subtle proliferation of a network of blood vessels that may penetrate into cancerous growths. Synchrotron radiation-based micro-computed tomography (SR-μCT) allows researchers to obtain 3D vascular tree at a micron-level resolution. However, most of them are qualitative description rather than quantitative analysis, which is still a world-wide challenge. There is a lack of software suitable for the automated quantitative analysis of 3D angiogenesis, especially for the analysis of microvasculature where the large size of each of the 3D image dataset needs to be processed within acceptable timeframes.

Informative visualization of the vasculature is a major step towards the aforementioned detection and assessment; however, objectivity, reliability and reproducibility cannot be achieved without quantitative information. This chapter is devoted to quantifying the vessel system in a number of information dimensions so that any changes can be readily detected by computers. Figure 10.1 illustrates the flow of the processing steps.

10.2 Ultra High-Resolution Imagery

10.2.1 Data Acquisition Protocol

SR-μCT can achieve a micron-level resolution, making microvasculature accessible with digital imaging. This makes accurate quantification of microvasculature changes feasible.

As part of a study to better evaluate this feasibility, as well as assess its potential in early diagnosis and treatment evaluation, multiple sets of liver and brain images of mice were captured at the Shanghai Synchrotron Radiation Facility (SSRF). Detail of the data acquisition protocol is given below.

The image datasets used in this chapter include cerebral and liver vasculatures of mice. The cerebral vessels were effectively perfused by Microfil. The cerebral vasculature images were acquired with the following experimental parameters: 13 keV

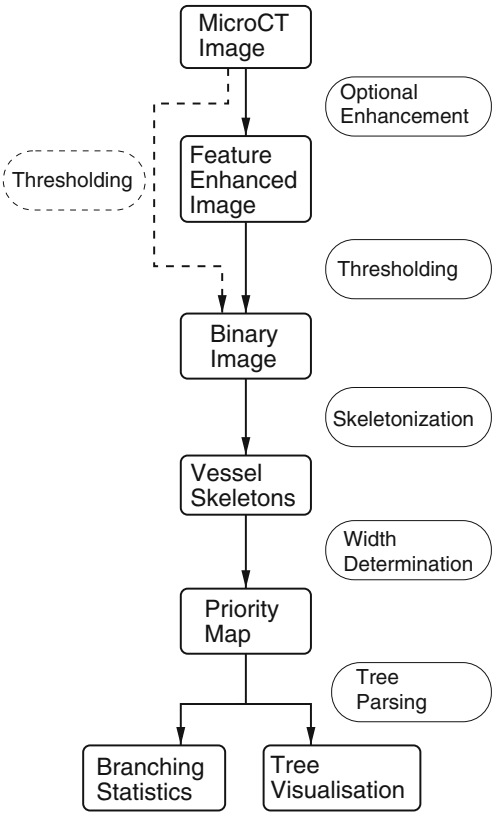


Fig. 10.1 Diagram of the processing flow

for the energy, 3.7 μm/pixel for the resolution, sample-to-detector distance (SDD) 30 cm. Mice bearing liver xenograft models were used to capture liver vasculatures. The liver orthotopic tumor xenograft models were established by implantation of HCCLM3-RFP tumor cells. The liver blood vessels were Heparinized saline perfused. There was no contrast agent used in the imaging. The experimental parameters were: 15 keV for the energy, 9 μm/pixel for the resolution, and 1 m for the SDD.

10.2.2 Significance of High Image Resolution

One of the advantages of having an ultra high resolution for biomedical purposes is that minute details, such as those significant to detecting or quantifying neovascularization, become considerably more accessible than otherwise. To illustrate this, a 3D visualization of the vasculature of a mouse brain, as captured at the SSRF, is presented in Fig. 10.2.

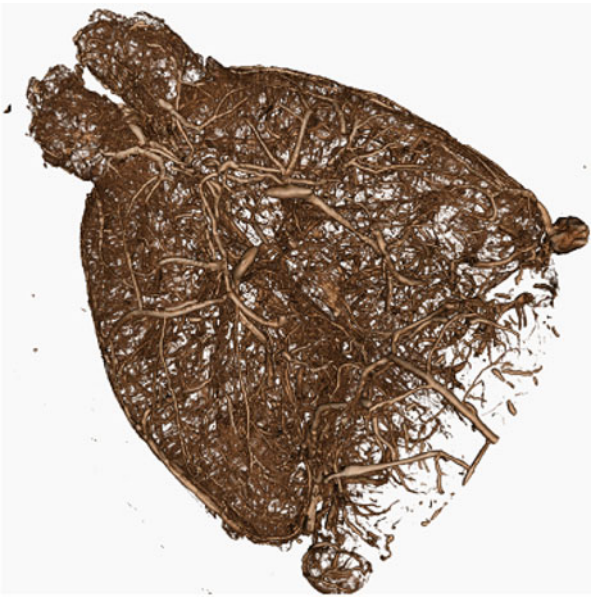


Fig. 10.2 An example of accessible details in ultra high-resolution imagery: a 3D visualization of the vasculature of a partial mouse brain captured at SSRF

10.3 Methods and Algorithms

10.3.1 Noise Levels and Linear-Structure Enhancement

The ultra high-resolution images that synchrotron radiation is capable of producing carry undesirable side-effects. Most noticeably, these include ring-like artifacts and high-level imaging noise, particularly those that coincide with, or are in proximity to, the inherent structure of interest (which are, in our case, the vessels).

Notwithstanding the more recent variants, the original vessel likelihood measure (referred to as vesselness) as proposed by Frangi et al.'s seminal work [5] is still widely used for locating vessels. Our current linear structure enhancement is based on the method described in [5]. Our work on 2D linear feature enhancement [20] will be extended to 3D in future work. One example of our test results is shown in Fig. 10.3.

In the pre-processing stage, the binary images were obtained by thresholding, where an automatic threshold selection approach that minimizes intra-class variance, generally known as Otsu's method [15,19], was employed for the contrast-enhanced cerebral images. Manual adjustments were needed for low-contrast images; however, it is hoped that when the above-mentioned enhancement becomes fully functional in 3D, the pre-processing stage will also become automated on all types of images.

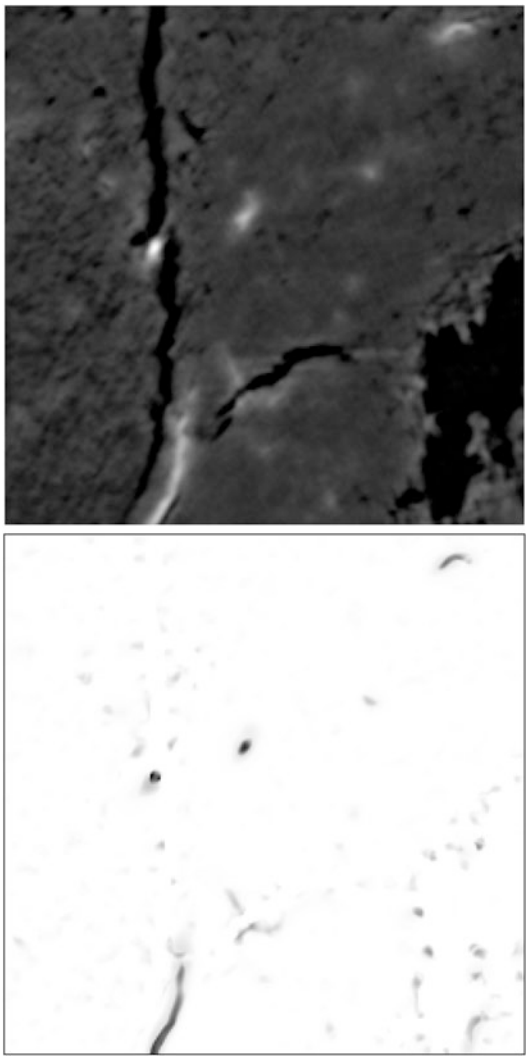


Fig. 10.3 Sub-image of a single synchrotron image slice before (*above*) and after (*below*) the enhancement (the intensity has been inverted for the enhanced image for display purposes)

10.3.2 Dimensionality Reduction by Skeletonization

In the analysis of vessel systems, the need to produce a hierarchical representation is ever present [12]. As an approach to representing and characterizing the linearity of vessels, skeletonization is typically required as a dimensionality reduction measure.

In general, skeletonization approaches on binary images can be discussed in two broad categories, divided along the line of whether or not image-weighted distances play a primary role in the process. In the category where such distances are used, skeletons are classically computed as part of the Voronoi diagrams [6]. Alternatively, they can be regarded as comprising of singular points (i.e., discontinuities in the first order derivatives) or regional maxima in a distance map [2, 16], where the image-weighted distance measure can be Euclidean or otherwise (for speed considerations a Chamfer metric is often employed [14]). A third approach is to administer simulated wave propagation with the speed dependent on image information [1, 18]. The propagating waves can be initiated either inside [13] or outside the structure of interest. If started outside, this operation is also referred to as the grass fire transform (GFT) [3, 11].

Where distances are either not being used, or only playing a supporting role, the approaches are generally called topological thinning, referring to the process in which voxels or pixels not affecting the topology of the underlying structure are iteratively removed [4], resulting in a morphologically thinned object of the original binary image. The decision as to whether or not to remove a voxel is typically conditioned upon a set of criteria, such as that the voxel is on the boundary of the object under consideration, that its removal will not alter the topology of either the object or the background, and it is non-terminal so that its deletion will not shorten the linear structure. Algorithms have been proposed to accelerate the process and improve the symmetricity of the process. These include the ordering of the voxels by their distances to the background, and the octree algorithm used to retrieve and organize a complete set of boundary voxels.

While demonstrating superior performance in terms of speed, distance map based approaches do not preserve any topological information in the process. The end results are typically a set of locations, without explicit topological information. However, a skeleton generally needs to incorporate topological information, in addition to location information. Therefore, the possibility of introducing errors in a post-processing stage, while attempting to recover the topological information (e.g., by linking voxels in proximity) cannot be discounted.

For guaranteed faithfulness down to the voxel level to the original topology of vasculature, a procedure based on the topological thinning [9, 10] was chosen, which essentially comprises conditional erosions. The actual implementation performs three essential tests at each iteration for each point under consideration: whether the point is lying on the surface of an object, whether it is a simple point, and whether it is non-terminal. The first criterion is the most crucial to the efficiency of the entire algorithm [7, 10]. That is, how many boundary points can be retrieved on each application of the greedy retrieval step. The implementation [7, 8] is one based on a kernel, and has been demonstrated to discover all surface points upon each application [10].

10.3.3 Quantitative Understanding of the Vasculature Hierarchy

The quantitative understanding module in the pipeline described here performs derivation of a vessel hierarchy. The skeleton of the blood vessels shown in Fig. 10.4 is split into segments and then a tree analysis is conducted. This analysis is based on the tree structure of vessel systems, as conceptually illustrated in Fig. 10.5.

This module is a modification and 3D extension of a previously published work dedicated to neurite analysis in 2D [21–23]. Vessel-segment trees are constructed by a process akin to region growing on a graph, or a marker controlled morphological

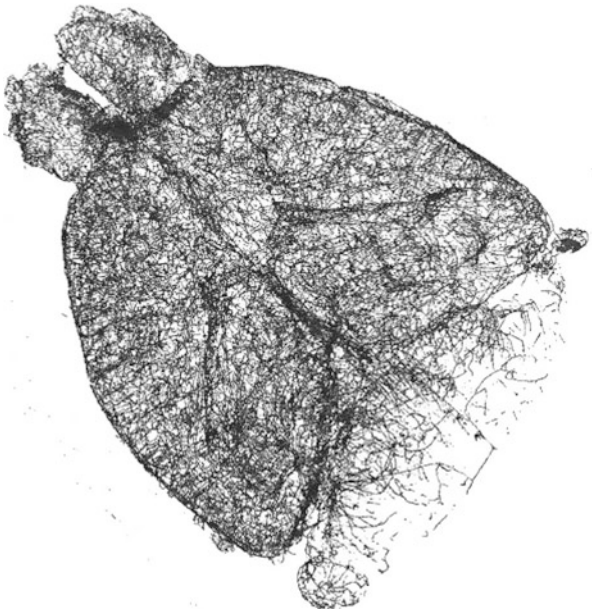


Fig. 10.4 A 3D visualization of the skeletonized cerebral vasculature (as shown in Fig. 10.2) from the partial mouse brain captured at SSRF

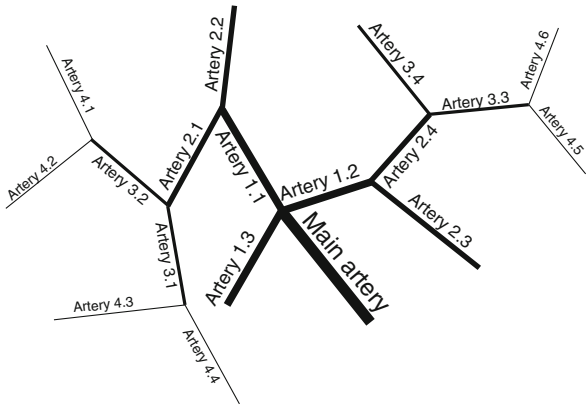


Fig. 10.5 A conceptual illustration of the hierarchical nature of a vessel system, with a main artery in the centre of the graph

watershed transform, where the edges are vessel bifurcations and the nodes are the skeleton segments. This process begins with segments that touch the main entry arteries, which can be identified either by a manual click or by a morphological process that searches for the thickest vessel within a designated volumetric area.

To delineate 3D vessel segments, which constitute the tree branches and on which the average thickness is measured, all the junctions and bifurcations are removed prior to the tree-structure parsing. Furthermore, very short segments, deemed insignificant, are also deleted.

As with the marker controlled morphological watershed transform, a priority map is needed to determine the order for the 'flooding' process. In the current context, each segment is given a priority value that is proportional to the average thickness of the segment. That is, the thicker the segment, the higher its priority. Starting at the main entry point (and, where necessary, restarted from the locations of the deleted junction branch points, which are analogous to the seeds in the marker-based watershed transform), the tree progressively grows primary, secondary, tertiary, and other levels of branches. Branches at each level of the hierarchy are assigned the same, unique label. Following the above parsing of the tree hierarchy, statistics in terms of quantitative properties such as the lengths and widths are produced and compiled into a table.

10.4 Results

In this section, we present the results of the quantification of vessel systems. Figure 10.6 demonstrates the segmented liver vasculature of a mouse, and Fig. 10.7 shows the segmented, labeled cerebral vascular trees of a mouse. Tables 10.1 and 10.2 show the quantitative analysis results for the liver and cerebral vascular trees, respectively.

Tables 10.1 and 10.2 include both an image-wide summary and vessel tree-based measurements. The image-wide summary, given at the top part of each table, includes the following statistics:

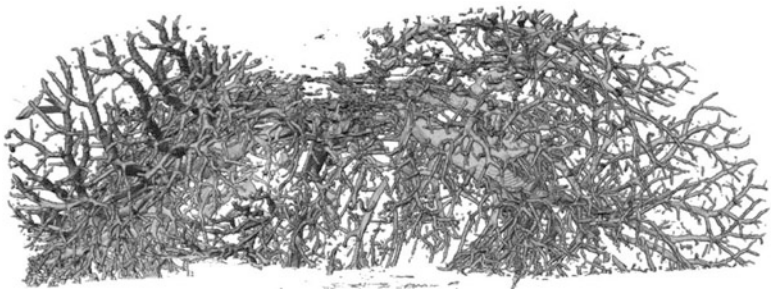


Fig. 10.6 3D visualisation of the segmented liver vasculature of a mouse

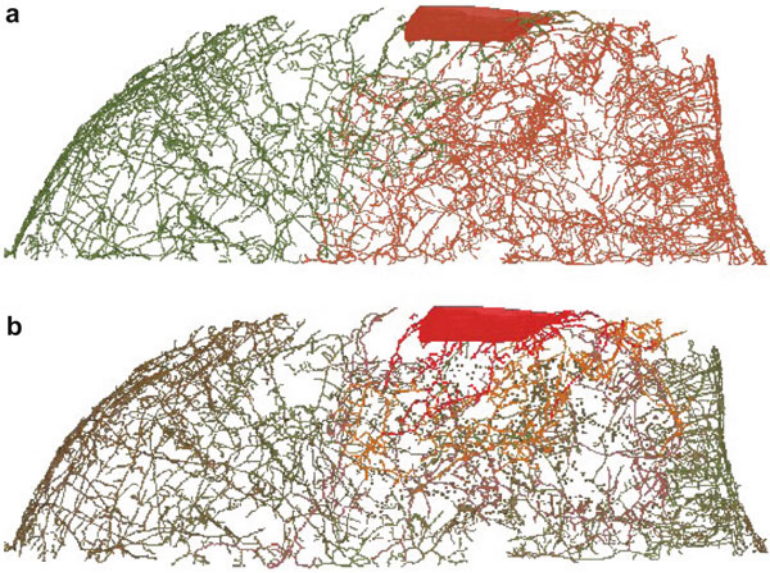


Fig. 10.7 Segmented and labeled cerebral vasculature of a mouse. (**a**) Labeled vasculature tree; (**b**) Labeled branch layer structure

1. Total length – the total length of all blood vessels (in voxels)
2. Total number of vessel segments – a segment is a linear structure between branching/junction points
3. Length of the longest vessel from marker – the length (in voxels) of the longest vessel from the marker to the end of the most distant segment
4. Total number of extreme branches – the number of terminating vessel segments
5. Total number of branch points – the number of points where a vessel structure splits into two or more branches

The vessel-tree based measurements include:

1. Tree label – sequential number of all vessel trees from the marker in the vascular structure
2. Total length – sum of the length of all the vessel branches of the tree
3. Length of longest vessel from marker – the length (in voxels) of the longest path from the marker to the end of the most distant vessel segment
4. Max branch layer – the highest level of branching for the tree
5. Mean branch layer – the mean level of branching for the tree
6. Number of branch points – the number of points where a vessel structure splits into two or more branches
7. Number of vessel segments – the number of segments where a segment is a linear structure between two neighbouring branching points.
8. Number of extreme branches – the number of terminating vessel segments

Table 10.1 Statistics of the mouse liver vasculature

Total length	Total number of vessel segments	Length of longest vessel from marker	Total number of extreme branches	Total number of branch points
19,338.1	4,204	285.30	1,737	932

Tree label	Total length	Length of longest vessel from marker	Max branch layer	Mean branch layer	Number of branch points	Number of vessel segments	Number of extreme branches
1	4.08	4.08	1	1	0	1	1
2	16.42	16.42	1	1	0	2	1
3	13.99	13.99	1	1	0	1	1
4	15.14	15.14	1	1	0	1	1
5	9.06	9.06	1	1	0	1	1
6	2.23	2.23	1	1	0	1	1
7	4,498.07	285.30	16	9.64	182	1,248	437
8	5.49	5.49	1	1	0	1	1
9	30.85	14.82	2	1.71	1	7	4
10	20.98	20.98	1	1	0	1	1
11	48.92	25.01	3	2	3	12	6
12	1,565.71	123.54	9	5.79	72	227	121
13	11,764.10	206.93	18	7.90	606	2,498	1,060
14	3	3	1	1	0	1	1
15	5.26	5.26	1	1	0	1	1
16	1,073.64	111.53	9	4.86	56	160	81
17	261.05	87.93	5	2.39	12	41	18

The total length of all branches of a tree (the second column), together with the statistics of the numbers of branch layers – given in the last five columns (the fourth to the eighth) of Table 10.1, have particular significance in detecting tumor growth. In the case presented here, the vessels belonging to the tree numbered 13 have a total of 11,764.10 mm in length, 18 maximum branch layers and 7.90 average layers. These make it the most note worthy amongst the 17 vessel trees, ranked just above the similarly complex tree number 7. In Table 10.2, the most significant trees are those numbered 6 and 34.

While it may not always be achievable to accurately register longitudinal images of the same patient (or experimental subject), the presence of such structurally riches sub-systems make direct comparisons feasible. Such comparison between longitudinal samples are crucial to detecting any such changes and, by extension, the possibility of tumor growths.

Table 10.2 Quantitative analysis results of the mouse cerebral vasculature

Total length	Total number of vessel segments	Length of longest vessel from marker	Total number of extreme branches	Total number of branch points
79,639.5	7,255	2,337.88	2,293	2,421

Tree label	Total length	Length of longest vessel from marker	Max branch layer	Mean branch layer	Number of branch points	Number of vessel segments	Number of extreme branches
1	65.10	65.10	1	1	0	1	1
2	31.77	31.77	1	1	0	1	1
3	145.59	71.98	3	2.28	4	28	11
4	46.86	46.86	1	1	0	1	1
5	30.25	30.25	1	1	0	1	1
6	47,615.20	2,007.31	39	21.23	1,449	4,317	1,377
7	53.71	53.71	1	1	0	1	1
8	188.78	66.80	3	2.35	4	28	9
9	225.00	132.52	6	3.47	8	19	6
10	36.85	26.02	2	1.60	1	5	2
11	42.22	26.36	3	1.87	2	8	4
12	218.34	105.71	3	1.86	8	23	9
13	55.97	28.95	3	1.88	2	9	4
14	3	3	1	1	0	1	1
15	5.29	5.29	1	1	0	1	1
16	78.85	71.05	2	1.25	1	4	2
17	6.81	6.81	1	1	0	1	1
18	158.90	46.50	6	3.5	8	22	4
19	98.27	45.68	2	1.5	2	6	1
20	44.54	38.19	2	1.25	1	4	1
21	7.04	7.04	1	1	0	1	1
22	120.43	34.31	4	2.18	4	22	8
23	39.08	39.08	1	1	0	1	1
24	3	3	1	1	0	1	1
25	24.49	24.49	1	1	0	2	0
26	26.91	26.91	1	1	0	1	1
27	98.60	98.60	1	1	0	1	1
28	65.27	54.05	2	1.25	1	4	1
29	38.93	38.93	1	1	0	1	1
30	41.38	41.38	1	1	0	1	1
31	134.36	121.09	2	1.4	2	5	2
32	331.26	170.67	3	2.02	10	37	15
33	40.61	37.10	2	1.33	1	3	2
34	29,428.60	2,337.88	44	26.96	912	2,689	816
35	24.41	24.41	1	1	0	1	1
36	36.20	25.42	2	1.33	1	3	2
37	27.31	27.31	1	1	0	1	1

Discussions and Conclusions

An approach has been presented to bringing new insights into the quantitative assessment of angiogenesis, and to analysing and detecting tumour angiogenesis at its early stage that could lead to tumor malignancies. Vascular trees are highly branching, asymmetric and non-homogeneous structures comprised of tapering vessels. In this study, the quantification of branching structure concerns the extracting of the statistical and geometric features of individual vascular segments. Some existing studies treated the structure of vascular tree as a global entity, lacking the branching structure analysis of local vessel segments such as vessel bifurcations. Besides reporting the global features of a vascular tree, such as vessel density, total length and diameter, the detailed branching structure reported in this chapter, including number of branching layers, number of branching points, number of vessel segments etc. will help quantify the subtle bifurcations. This is essential in detecting the microvascular tree changes which may lead to tumor formation and growth.

There are several challenges in this study, including the quantification of branching structure of dense microvascular trees, and processing large 3D image datasets. This chapter addressed the quantification of the branching structure. The microvascular tree analysis is a compute-intensive task requiring high performance computing. In the future, we will aim to resolve the problems associated with the processing of large 3D image datasets.

Acknowledgements This research is supported by CSIRO – Chinese Academy of Sciences Collaborative Research Fund and CSIRO Computational and Simulation Sciences Transformational Capability Platform (CSS TCP). This work benefited from the use of the Insight Segmentation and Registration Toolkit (ITK) and the Visualization Toolkit (VTK), open source software developed as an initiative of the U.S. National Library of Medicine and available at www.itk.org.

References

1. D. Adalsteinsson, J.A. Sethian, A fast level set method for propagating interfaces. J. Comput. Phys. **118**, 269–277 (1995)
2. C. Arcelli, G.S. di Baja, L. Serino, Distance-driven skeletonization in voxel images. IEEE Trans. Pattern Anal. Mach. Intell. **33**(4), 709–720 (2011)
3. H. Blum, Biological shape and visual science (part I). J. Theor. Biol. **38**, 205–287 (1973)
4. A.V. Evako, Characterizations of simple points, simple edges and simple cliques of digital spaces: one method of topology-preserving transformations of digital spaces by deleting simple points and edges. Graph. Models **73**(1), 1–9 (2011)
5. A.F. Frangi, W.J. Niessen, K.L. Vincken, M.A. Viergever, Multiscale vessel multiscale vessel enhancement filtering, in *Medical Image Computing and Computer-Assisted Intervention*, Cambridge (Springer, 1998), pp. 130–137
6. A. Franz, Voronoi diagrams: a survey of a fundamental geometric data structure. ACM Comput. Surv. **23**(3), 345–405 (1991)
7. H. Homann, Implementation of a 3D thinning algorithm. Insight J. July-Dec 2007

8. L. Ibanez, W. Schroeder, L. Ng, J. Cates, *The ITK Software Guide: The Insight Segmentation and Registration Toolkit* (Kitware Inc., Clifton Park, 2003)
9. P.P. Jonker, Skeletons in N dimensions using shape primitives. Pattern Recognit. Lett. **33**(6), 677–686 (2002)
10. T.C. Lee, R.L. Kashyap, C.N. Chu, Building skeleton models via 3-D medial surface/axis thinning algorithms. Comput. Vis. Graph. Image Process. **56**(6), 462–478 (1994)
11. F. Leymarie, M. Levine, Simulating the grassfire transform using an active contour model. IEEE Trans. Pattern Anal. Mach. Intell. **14**, 56–75 (1992)
12. R. Li, S. Brown, L. Wilson, J. Young, S. Luo, Progressively refined patient-specific vessel system models from generic representations, in *Proceedings of the Sixth Digital Image Computing: Techniques and Applications*, Melbourne, pp. 184–189 (2002)
13. R. Li, S. Ourselin, Combining front propagation with shape knowledge for accurate curvilinear modelling, in *Medical Image Computing and Computer-Assisted Intervention*, ed. by R. Ellis, T. Peters (Springer, Berlin/Heidelberg, 2003), pp. 66–74
14. R. Li, S. Ourselin, Toward consistently behaving deformable models for improved automation in image segmentation, in *Deformable Models – Biomedical and Clinical Applications*, ed. by A. Farag, J.S. Suri (Springer, New York, 2007), pp. 259–292
15. N. Otsu, A threshold selection method from gray-level histograms. IEEE Trans. Syst. Man Cybern. **1**, 62–66 (1979)
16. A. Rosenfeld, J.L. Pfaltz, Distance functions on digital pictures. Pattern Recognit. **1**, 33–61 (1968)
17. M. Rücker, M.W. Laschke, D. Junker, C. Carvalho, A. Schramm, R. Mülhaupt, N.C. Gellrush, M.D. Menger, Angiogenic and inflammatory response to biodegradable scaffolds in dorsal skinfold chambers of mice. Biomaterials **27**(29), 5027–5038 (2006)
18. J.A. Sethian, A fast marching level set method for monotonically advancing fronts. Proc. Natl. Acad. Sci. U.S.A. **93**(4), 1591–1595 (1996)
19. M. Sezgin, B. Sankur, Survey over image thresholding techniques and quantitative performance evaluation. J. Electron. Imaging **13**, 146–165 (2004)
20. R. Su, C. Sun, C. Zhang, T.D. Pham, A new method for linear feature and junction enhancement in 2D images based on morphological operation, oriented anisotropic Gaussian function and Hessian information. Pattern Recognit. **47**(10), 3193–3208 (2014)
21. C. Sun, P. Vallotton, Fast linear feature detection using multiple directional non-maximum suppression. J. Microsc. **234**(2), 147–157 (2009)
22. P. Vallotton, R. Lagerstrom, C. Sun, M. Buckley, D. Wang, M. De Silva, S.-S. Tan, J. Gunnersen, Automated analysis of neurite branching in cultured cortical neurons using HCA-Vision. Cytometry **71A**(10), 889–895 (2007)
23. D. Wang, R. Lagerstrom, C. Sun, L. Bischof, P. Vallotton, M. Götte, HCA-Vision: automated neurite outgrowth analysis. J. Biomol. Screen. **15**(9), 1165–1170 (2010)

Chapter 11
Cloud Based Toolbox for Image Analysis, Processing and Reconstruction Tasks

Tomasz Bednarz, Dadong Wang, Yulia Arzhaeva, Ryan Lagerstrom, Pascal Vallotton, Neil Burdett, Alex Khassapov, Piotr Szul, Shiping Chen, Changming Sun, Luke Domanski, Darren Thompson, Timur Gureyev, and John A. Taylor

Abstract This chapter describes a novel way of carrying out image analysis, reconstruction and processing tasks using cloud based service provided on the Australian National eResearch Collaboration Tools and Resources (NeCTAR) infrastructure. The toolbox allows users free access to a wide range of useful blocks of functionalities (imaging functions) that can be connected together in workflows allowing creation of even more complex algorithms that can be re-run on different data sets, shared with others or additionally adjusted. The functions given are in the area of cellular imaging, advanced X-ray image analysis, computed tomography and 3D medical imaging and visualisation. The service is currently available on the website www.cloudimaging.net.au.

T. Bednarz (✉) • N. Burdett
Digital Productivity Flagship, CSIRO, Brisbane, QLD, Australia
e-mail: tomasz.bednarz@csiro.au; neil.burdett@csiro.au

D. Wang • Y. Arzhaeva • R. Lagerstrom • P. Vallotton • P. Szul • S. Chen • C. Sun
Digital Productivity Flagship, CSIRO, Locked Bag 17, North Ryde, Sydney, NSW 1670, Australia
e-mail: dadong.wang@csiro.au; yulia.arzhaeva@csiro.au; ryan.lagerstrom@csiro.au; pascal.vallotton@csiro.au; piotr.szul@csiro.au; shiping.chen@csiro.au; changming.sun@csiro.au

A. Khassapov • D. Thompson
CSIRO IM&T, Melbourne, VIC, Australia
e-mail: alex.khassapov@csiro.au; darren.thompson@csiro.au

L. Domanski
CSIRO IM&T, Sydney, NSW, Australia
e-mail: luke.domanski@csiro.au

T. Gureyev
CSIRO Manufacturing Flagship, Melbourne, VIC, Australia
e-mail: tim.gureyev@csiro.au

J.A. Taylor
Digital Productivity Flagship, CSIRO, Canberra, ACT, Australia
e-mail: john.a.taylor@csiro.au

C. Sun et al. (eds.), *Signal and Image Analysis for Biomedical and Life Sciences*,
Advances in Experimental Medicine and Biology 823,
DOI 10.1007/978-3-319-10984-8_11

Keywords Cloud computing • Cloud based imaging • Image analysis • Image processing • Image reconstruction

11.1 Introduction

Imaging technologies, especially in the biomedical areas, have seen tremendous progress in the past decades. There are about 40,000 magnetic resonance imaging (MRI) scanners in operation worldwide providing image contrasts and resolution never achieved before. Computed tomography (CT), positron emission tomography (PET), and microscopy are other modalities becoming fundamental tools for many scientific areas [8, 20, 23]. CSIRO (www.csiro.au) is involved in many collaborative projects and clinical studies, such as the synchrotron based X-ray microscopy and micro-spectroscopy, providing exquisite details down to the sub-cellular structures resolution.

The primary goal of this work was to provide improved access to the existing biomedical image processing and analysis software packages to nation-wide research communities via remotely accessible user-interfaces, which is supported and carried out on the NeCTAR [21] supported cloud infrastructures (www.nectar.org.au). These software packages have been developed over the last 10–15 years by CSIRO scientists and software engineers, and they include:

- *HCA-Vision* [17]: developed for automating the process of quantifying cell features in microscopy images. It can reproducibly analyse complex cell morphologies. It has considerable value in particular for the pharmaceutical and neuroscience research community.
- *MILXView* [19]: 3D medical image analysis and visualisation platform increasingly popular with researchers and medical specialists working with MRI, PET and other types of medical images. A suite of functions has been developed for viewing and processing 3D and 4D medical data. Several advanced processing pipelines also exist, such as a fully automated brain morphometry estimate from MRI.
- *X-TRACT* [26]: developed for advanced X-ray image analysis and computed tomography currently in use on the Multi-modal Australian ScienceS Imaging and Visualisation Environment cluster (MASSIVE, see www.massive.org.au) at the Australian Synchrotron, the Australian National University (ANU, see www.anu.edu.au) and at the Shanghai Synchrotron (ssrf.sinap.ac.cn/english/) in China. *X-TRACT* implements a large number of conventional and advanced algorithms for 2D and 3D X-ray image reconstruction and simulation.

The service unifies those software packages for image analysis, processing and reconstruction. Each function is provided as a building block, which can be assembled together in a graphical workflow environment, using the Galaxy package [13–15, 22] (galaxyproject.org). By providing user-friendly access to cloud computing resources and new workflow-based interfaces, our solution enables the researchers to carry out many challenging image analysis and reconstruction tasks.

The main focus at this stage is on algorithms and solutions primarily for biomedical image processing and analysis; however, the tools being developed have no inherent technical restriction and can be extended to support a wide variety of fields of scientific research, such as [2, 7, 24].

11.2 Software Architecture

The cloud-base image analysis and processing toolbox [9] comprises a collection of physical and virtualized resources connected through networks, including the NeCTAR research cloud Infrastructure as a Service (IaaS), cloud enabled image analysis and processing Platform as a Service (PaaS), and CSIRO developed image analysis Software as a Service (SaaS). Figure 11.1 presents a high-level architectural view of the system.

11.2.1 Platform as a Service

The image analysis and processing platform (PaaS) represents the development and runtime environment where the image analysis and processing tools are executed. The platform also provides the basic management features of a single node and

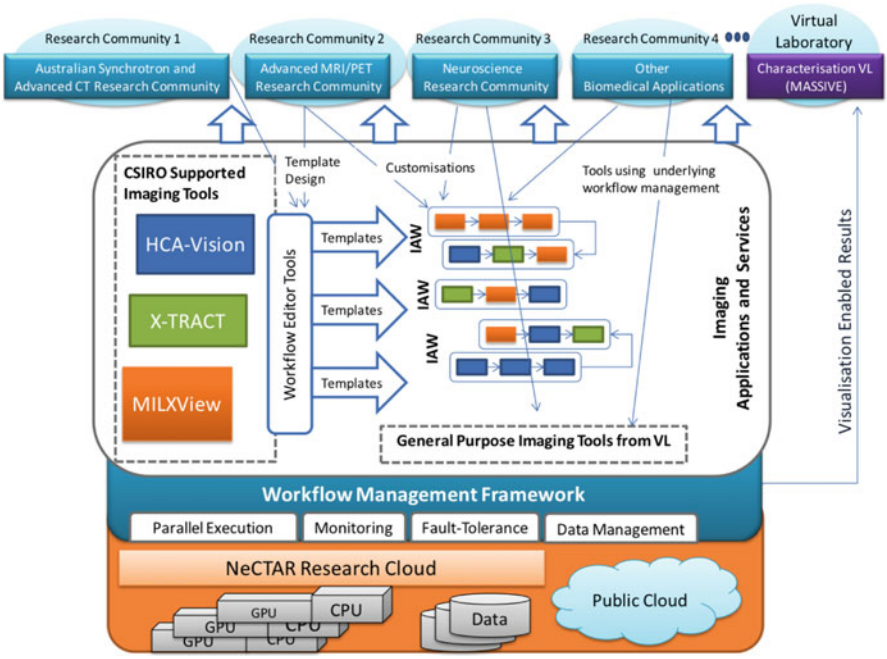


Fig. 11.1 System architecture (*VL* = Virtual Laboratory, *IAW* = Image Analysis Workflow)

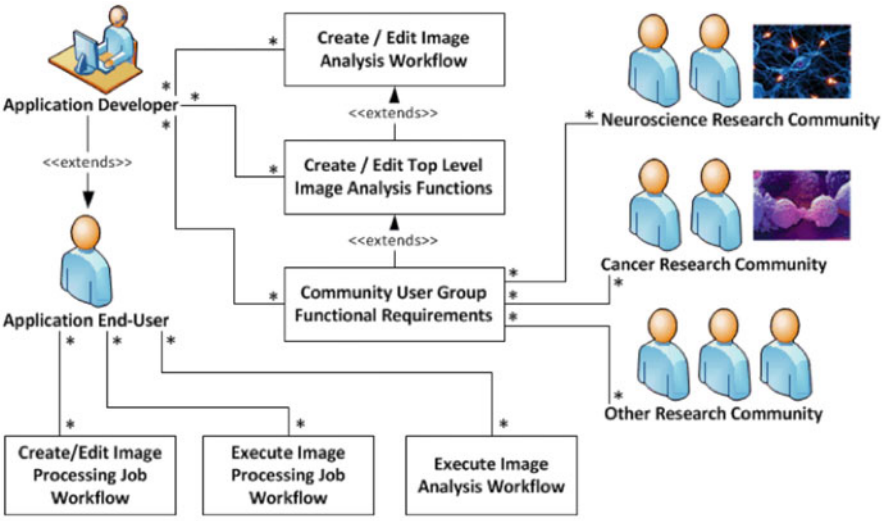


Fig. 11.2 Workflows use case

leverages all the other operations on the services that it is hosting. The platform encapsulates a layer of software and provides it as a service that can be used to build high level image analysis and reconstruction services. Figure 11.2 shows the workflows of a use case.

11.2.2 Software as a Service

The SaaS layer features three applications offered as a service on demand. A single instance of the tools extracted from each of the image analysis packages runs on the cloud and services multiple end users. The software packages include HCA-Vision [17], X-TRACT [26], and MILXView [19] that were mentioned earlier. There are no technical restrictions to run the service on other cloud infrastructures [3, 16, 18].

11.3 Launch of the Service

To facilitate end users' easy access to the NeCTAR research cloud, the cloud utilises the Australian Access Federation Registry AAF [1] to provide a web portal based single sign-on for users with the same login credentials that they use for login to their institutional networks (universities or research facilities).

To launch CSIRO Cloud-based Image Analysis and Processing Toolbox [5, 6, 11, 12], start any web browser and type the following web address: cloudimaging.net.au. Figure 11.3 shows the web page that pops up.

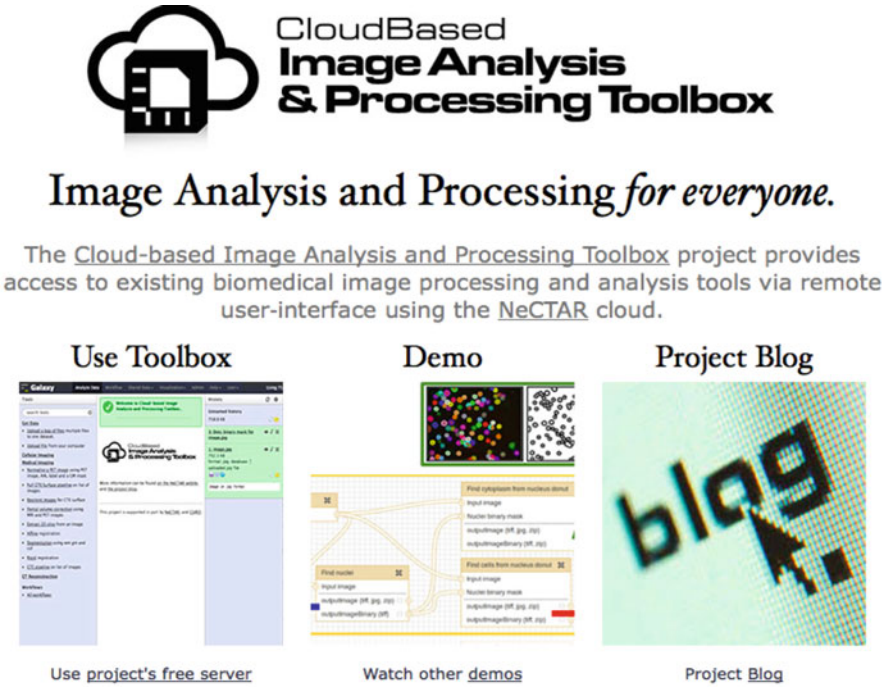


Fig. 11.3 Dasboard of the imaging service available under cloudimaging.net.au

Click the *Blog* [10] link (cloudimaging.blogspot.com.au), you will be able to access the blogs of the toolbox, and click *Demos* (www.youtube.com/user/CloudImaging), you will see videos showing how to use the toolbox. Click the "project's free server" web link (galaxy.cloudimaging.net.au) at the bottom left of the home page to bring up the Login page.

Type your email address and password which are the same as you used for logging into your institutional network, click *Login* to login to the cloud. The web portal will be shown in Fig. 11.4 and the user will have access to the following listed functions:

- *Get Data* – For a user to upload images, upload and merge multiple files into a single dataset, or split a multiple file dataset into standard files.
- *Cellular Imaging* – The cellular image analysis tools include automated solutions for cell image analysis. They include automated nucleus detection, cytoplasm detection, cell detection, dots and linear feature detection within a cell, retrieving statistical features of dots, lines, cytoplasm, cells etc.
- *Medical Imaging* – The medical image analysis toolbox comprises a suite of functions for processing and visualizing 3D and 4D medical images, such as image normalization, atlas registration, bias field correction, partial volume estimation, brain topology correction, cortex thickness estimation, cortical surface extraction, biomarker mapping on cortical surface, etc.

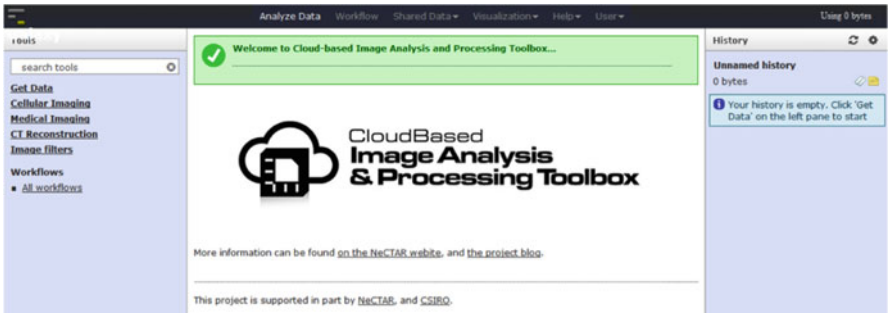


Fig. 11.4 Main homepage encapsulating all the functionalities

- *CT Reconstruction* – Tools for CT reconstruction include sinogram creation, ring artifact removal, dark current subtraction, flat field correction, positional drift correction, data normalization, filtered back-projection (FBP) parallel-beam CT reconstruction, Feldkamp-Davis-Kress (FDK) cone-beam CT reconstruction, automated detection of the centre of rotation in a CT scan, CT reconstruction filters, region of interests reconstruction, etc.
- *Image Filters* – This category contains image processing filters, including generic procedures and algorithms that are performed without a priori knowledge about the specific features of an image. Implemented filters include: Emboss, Sharpen, Min, Edge Enhance, Contour, Max, Invert, Median, Edge Enhance More, Detail, Blur, Smooth, Smooth More, Mode, and Find Edges.
- *Image Operations* – These tools include basic arithmetic and logical operations between a pair of images or between an image and a constant, performed pixel-wise.
- *Image Tools* – Tools include more general category and functions: Image Tileify, Image Resize, Image Rotate, and Image Filmify.

11.4 Supported Functions

Supported and available functions in the biomedical space are listed below. Table 11.1 lists the supported and available tools for cellular image analysis; Table 11.2 shows a list of medical imaging tools; and Table 11.3 lists available tools for CT reconstruction functionality.

Figure 11.5 depics applications of MILXView software for image registration (overlaying various image modalities) and capability for studying of atrophy pattern characteristics of diseases such as the Alzheimer's disease.

Figure 11.6 shows the capability of X-TRACT software [26] (which is part of the cloud imaging project): reconstructed insect (3D) and a sample input sinogram. The standalone software has a rich graphical user interface and tools for 2D image display and manipulation. X-TRACT can simulate a variety of phase-contrast imaging modes and can perform phase retrieval using a number

Table 11.1 List of cellular imaging tools available in the cloud-based image analysis and processing toolbox

Name of the tool	Description
Find cytoplasm from nucleus donut	Identify cytoplasm as a donut-like region around a nucleus
Detect dots	Detect dot-like structures in a 2D image
Create donut around a nucleus	Create a donut-like region around a cell nucleus
Objects statistics	Compute statistics of objects in a binary or labelled image
Detect nuclei from cytoplasm holes	Detect nuclei from absence of stain of cytoplasm
Detect lines	Detect linear features in an image
Arithmetic operations with two images	Add, subtract, multiply, and divide two images
Detect nuclei	Detect nuclei in a 2D microscope image
Detect cells	Detect cells with or without using a nucleus image as a mask
Overlay image with labelled mask	Create a colourful overlay image from an original image and a labelled mask
Statistics for lines	Compute statistics for lines in an image
Detect neurites	Detect neurites in a 2D microscope image
Label objects	Label objects in a binary image from 1 to N where N is the number of objects
Filter objects by morphology	Filter objects in an image by their morphological properties
De-clump touching objects	Separate any merged objects such as merged nuclei or cells
Statistics for dots	Compute statistics for dots in the image
Detect neurons	Detect neurons from a neurite outgrowth image with or without using a nucleus image as a mask
Logical operations with two images	Logical operations with two images such as AND, OR, and XOR
Extract image channel	Extract red, green or blue channel from a color image

of different methods. The program can also simulate polychromatic tomographic projections in parallel and cone-beam modes, where the sample is represented as a stack of 2D slices with a known spatial distribution of real and imaginary parts of the refractive index, or a distribution of known material components. X-TRACT contains algorithms for parallel-beam and cone-beam CT reconstruction.

11.5 Workflows

The image analysis and processing platform represents the runtime environment where above listed image analysis and processing tools are executed. The functions could be lined up and connected together to create execution pipeline, as shown in Fig. 11.7.

Table 11.2 List of medical imaging tools available in the cloud-based image analysis and processing toolbox

Name of the tool	Description
Extract 2D slices	Extract 2D slices from a 3D image so a user can view the data they have uploaded
Image registration	Aligning 2 datasets, from different co-ordinate systems, together
Image segmentation	Segment the image into multiple areas for further analysis
Cortical thickness estimation (CTE)	Quantify loss of grey matter in cortex
CTE surface	Used to transfer CTE results to a common template mesh for comparison with populations
Reorient images	Reorient images in the space
Standard uptake value ratio (SUVR)	Normalise intensity of a PET image
Partial volume correction (PVC)	De-blurring of images

Table 11.3 List of CT reconstruction tools available in the cloud–based image analysis and processing toolbox

Name of the tool	Description
CT reconstruction	Create a slice from a sinogram. X-ray projection data must first be converted into sinograms before CT reconstruction can be carried out. Each sinogram contains data from a single row of detector pixels for each illuminating angles. This data is sufficient for the reconstruction of a single axial slice (at least, in parallel-beam geometry). See Fig. 11.6
Center of rotation	Automated calculation of the centre of sample rotation in a CT scan from experimental X-ray projections, sinograms or reconstructed axial slices
Sinogram creation and preprocessing	Create sinograms from projections
FTP functions	Dataset must be copied to FTP folder before using FTP client for downloading

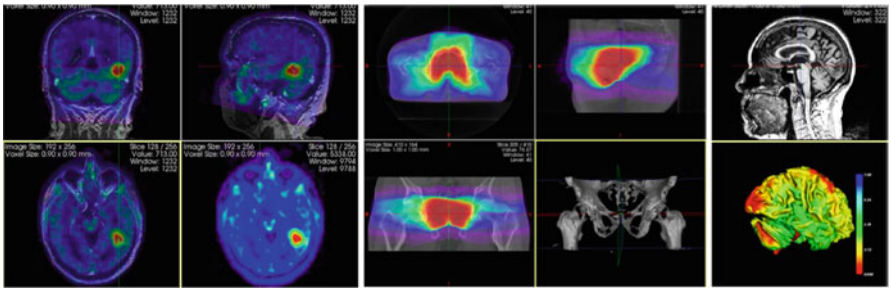


Fig. 11.5 Applications of MILXView: (*left*) Brain tumor – PET scan and MRI overlaid; (*middle*) CT scan of a prostate of a patient overlaid with radiation dose; (*right*) Generated 3D view of a brain allowing study of atrophy pattern characteristics of diseases such as the Alzheimer's disease (Reprinted with permission from [4]. Copyright 2013, AIP Publishing LLC)

Fig. 11.6 X-TRACT capabilities. *Top*: insect, reconstruction and rendering by Sherry Mayo of CSIRO. *Bottom*: sample input sinogram

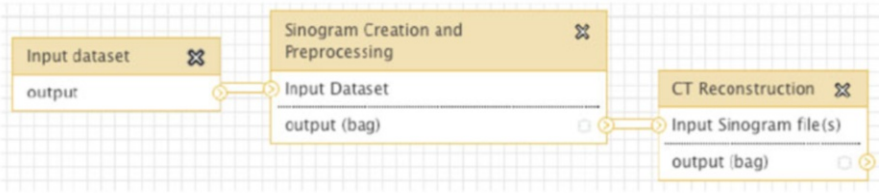


Fig. 11.7 Sample CT reconstruction workflow

The workflows could then be stored in the cloud and shared with others. Parameters can also be tuned. They can be reused in the future and be re-run on different datasets. Figure 11.8 shows an example workflow that counts primary neurite branches. Figure 11.9 shows subset of the input image and result of tool execution. The toolbox provides different viewers, allowing displaying outcomes of the processing steps in 2D or 3D (see Fig. 11.10).

11.6 Use Case: Detecting Neurites

Cellular imaging attracts the interests of both pharmaceutical industry and academia. Researchers can use the cellular image analysis tools to carry out high content analysis for their biomedical research. The image analysis tools provided in the cloud-based services can help them to conduct automated measurement of cell morphology and analysis of cellular responses in individual cells treated with different chemical compounds.

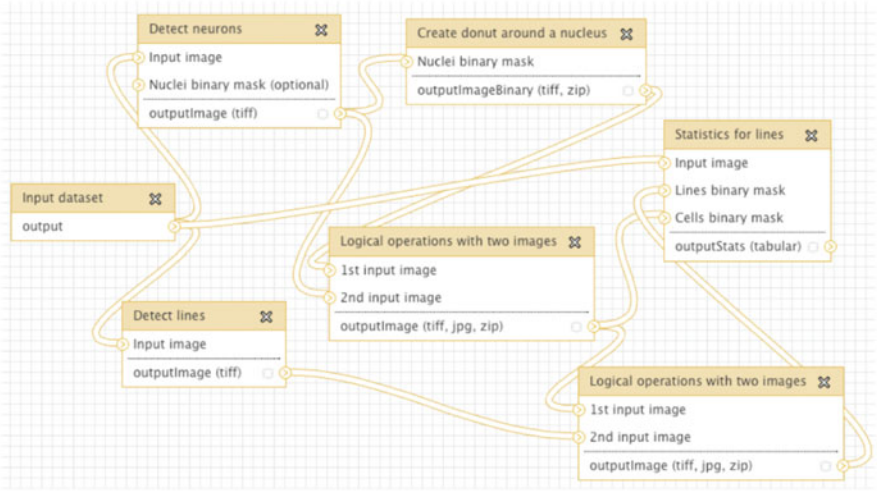


Fig. 11.8 Workflow for counting primary neurite branches

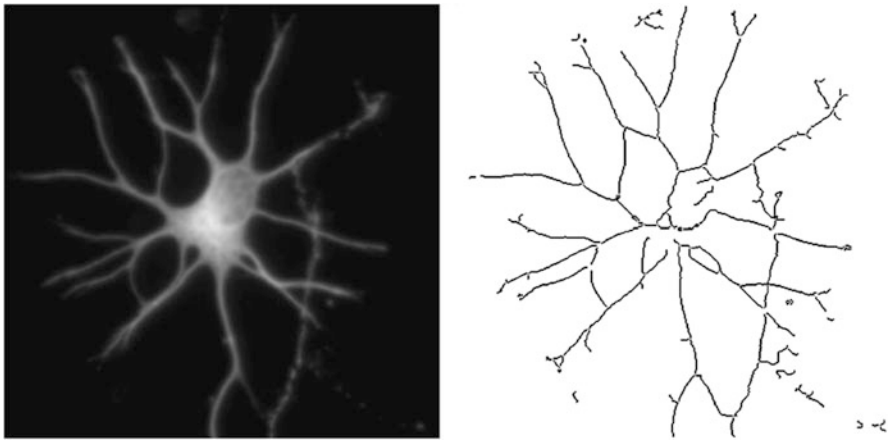


Fig. 11.9 *Left*: subset of the input image; *right*: lines with binary mask for neurite output

Typical input dataset looks like the one in Fig. 11.11. We want to automatically detect neurites in the picture. We carry out the detection in just a few steps using available functions:

- Upload data: Tools Panel → Get Data → Upload File (Fig. 11.11).
- Detect neurites: Cellular Imaging → Detect Neurites → Set Parameters → Execute (Figs. 11.12 and 11.13).
- Display the neurites on top of the input image: Cellular Imaging → Overlay Image with Labelled Mask → Set Parameters → Execute (Fig. 11.14).

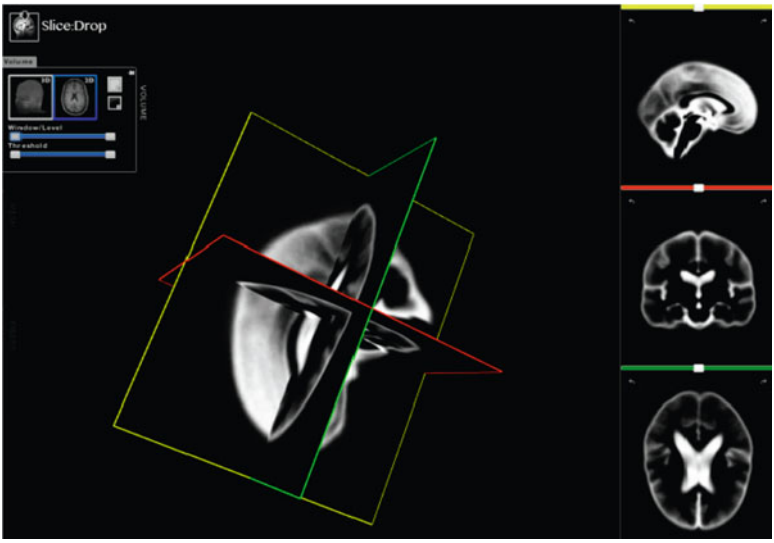


Fig. 11.10 Realtime 3D visualisation capability (WebGL based viewer [25])

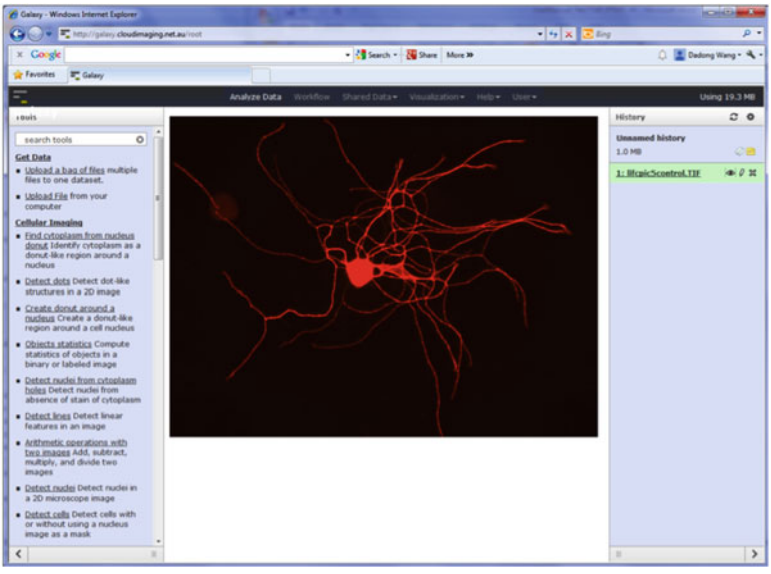


Fig. 11.11 Neurite outgrowth image uploaded to the cloud. Uploaded dataset is seen on the right-hand panel. When a user clicks on the *Eye* icon, the image is displayed in the center view as seen in this figure

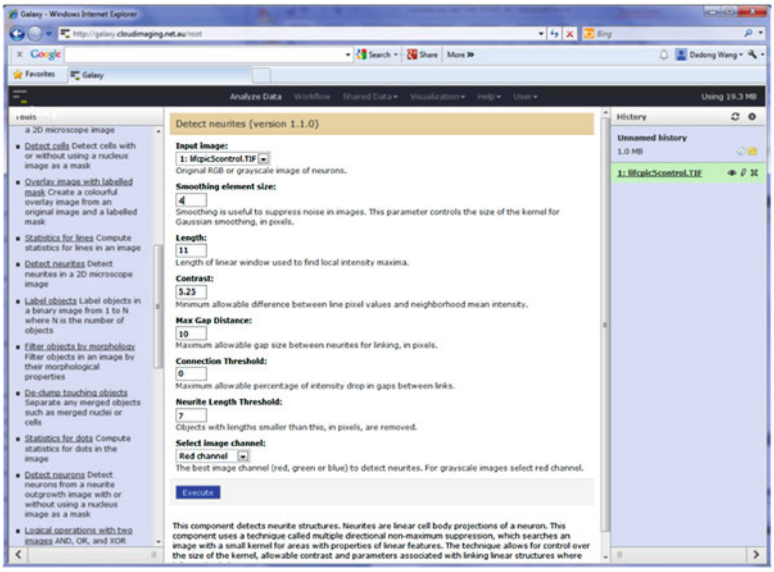


Fig. 11.12 Neurite detection parameters. The user can specify different parameters in the forms provided to almost every function. After parameters are set, a user clicks on *Execute* to run the function on the specified dataset in the cloud

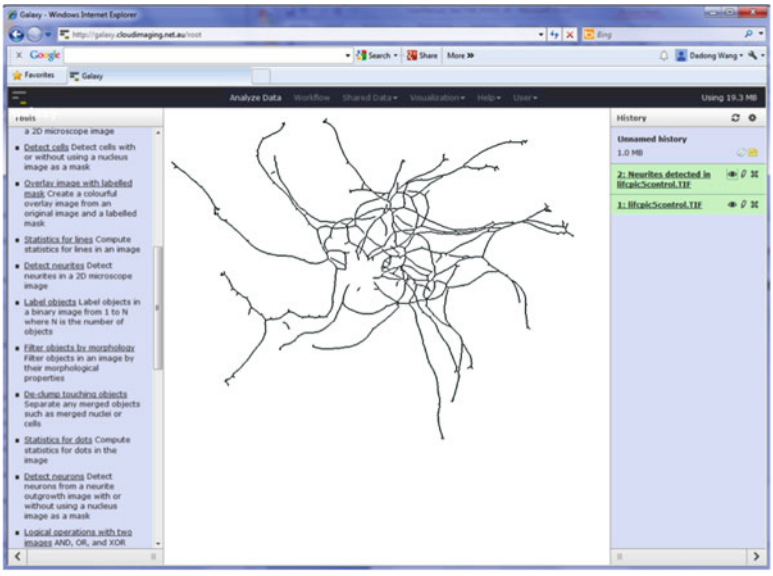


Fig. 11.13 Neurites detected. After the functions are executed, the user can display the output of the operation by clicking on the *Eye* icon

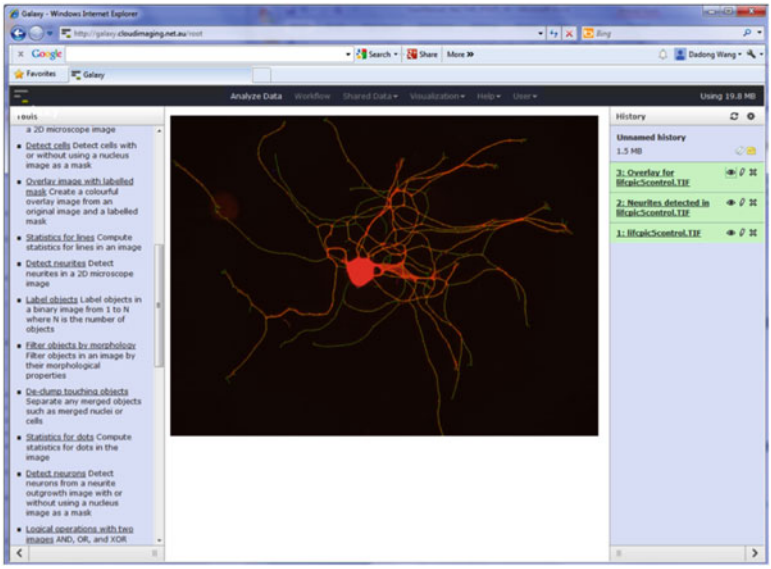


Fig. 11.14 Overlaid image showing the detected neurites on top of the neurite outgrowth image (for checking whether the result is properly aligned with the input image)

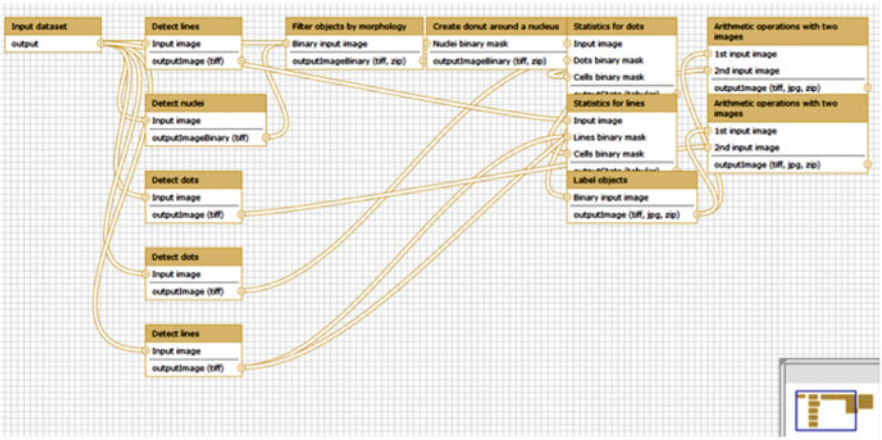


Fig. 11.15 Workflow created from astrocyte analysis history (a user can create history automatically – it is extracted using operations he or she executed earlier one by one)

The operations we carry out can be saved as a workflow. This workflow can be later used to recreate the operations on different datasets, or can be easily shared with others. Figure 11.15 shows a sample workflow created from astrocyte analysis history.

Conclusions
As datasets increase in size and complexity there is a significant burden placed upon existing hardware, infrastructure, and maintenance. Institutions are forced to spend heavily to upgrade and maintain hardware in order to increase productivity, by reducing processing times and increasing storage capacity. In this chapter we present a generic cloud framework that allows users access to complex pipelines such as those needed to process biomedical imaging data in a cloud environment.

The cloud can be hosted by a remote organisation, such as NeCTAR, thus relieving the institution of its expensive hardware and maintenance, and providing a possibly infinite number of machines for processing and data storage. It promotes collaboration, as users are able to access the tools and data from anyway in the world given the appropriate security credentials and an internet connection.

This work describes in general our imaging tools that are now available on the NeCTAR infrastructure under the following link: www.cloudimaging.net.au. Please refer to the user manual (goo.gl/99mEPV) for detailed description of the functionalities.

Acknowledgements The authors would like to acknowledge NeCTAR and CSIRO for funding for the project RT035 Cloud Based Image Analysis and Processing Toolbox.

References

1. AAF Website, (2013), see www.aaf.edu.au
2. M. Almeer, Cloud hadoop map reduce for remote sensing image analysis. J. Emerg. Trends Comput. Inf. Sci. **3**, 637–644 (2012)
3. Amazon, Amazon Web Services (AWS) (2013), see aws.amazon.com
4. T. Bednarz, P. Szul, Y. Arzhaeva, D. Wang, N. Budett, A. Khassapov, S. Chen, P. Vallotton, R. Lagerstrom, T. Gureyev, J. Taylor, Biomedical image analysis and processing in clouds, in *International Symposium on Computational Models for Life Sciences*, Sydney, vol. 1559, Nov 2013, pp.77–79
5. T. Bednarz, D. Wang, Y. Arzhaeva, P. Szul, S. Chen, A. Khassapov, N. Budett, T. Gureyev, J. Taylor, Cloud-based image analysis and processing toolbox for biomedical applications, in *8th IEEE International Conference on eScience*, Chicago, Oct 2012, pp. 8–12
6. T. Bednarz, D. Wang, Y. Arzhaeva, P. Szul, S. Chen, A. Khassapov, N. Budett, T. Gureyev, J. Taylor, Cloud-based workflows tools supporting medical imaging, in *International Conference on the Use of Computers in Radiation Therapy ICCR 2013*, Melbourne, May 2013, pp. 6–9
7. G. Berriman, E. Deelman, P. Groth, G. Juve, The application of cloud computing to the creation of image mosaic and management of their provenance, in *SPIE Conference 7740: Software and Cyberinfrastructure for Astronomy*, San Diego, ed. by N. Radziwill, A. Bridger, 2010
8. Characterisation Virtual Laboratory, (2013), see www.massive.org.au/special-projects/characterisation-virtual-laboratory

9. S. Chen, T. Bednarz, P. Szul, D. Wang, Y. Arzhaeva, N. Burdett, A. Khassapov, Galaxy + Hadoop: toward a collaborative and scalable image processing toolbox in cloud, in *3rd International Workshop on Cloud Computing and Scientific Applications (CCSA)*, Berlin, 2013
10. Cloud Based Image Analysis and Processing Toolbox Blog, see cloudimaging.blogspot.com.au
11. Cloud Based Image Analysis and Processing Toolbox Dashboard, see cloudimaging.net.au
12. Cloud Based Image Analysis and Processing Toolbox User Manual, see goo.gl/99mEPV
13. CloudMan, see usecloudman.org
14. Galaxy Website, (2012), see usecloudman.org
15. Genomics Virtual Laboratory, (2013), see genomicsvirtuallab.wordpress.com
16. Google App Engine, (2013), see code.google.com/appengine
17. HCA-Vision, see www.hca-vision.com
18. Microsoft (2012), Windows Azure, see www.microsoft.com/windowazure
19. MILXView 3D, see http://research.ict.csiro.au/software/milxview
20. Multi-modal Australian ScienceS Imaging and Visualisation Environment (MASSIVE), see www.massive.org.au
21. NeCTAR Official Website, see www.nectar.org.au
22. OpenStack Website, see www.openstack.org
23. Shanghai Synchrotron, see ssrf.sinap.ac.cn
24. SIMAGIS, Cloud Software for Microscopy and Image Analysis (2013), live.simagis.com/home
25. Slice:Drop website (2013), slicedrop.com
26. X-TRACT, see http://www.ts-imaging.net

Chapter 12
Pollen Image Classification Using the Classifynder System: Algorithm Comparison and a Case Study on New Zealand Honey

Ryan Lagerstrom, Katherine Holt, Yulia Arzhaeva, Leanne Bischof, Simon Haberle, Felicitas Hopf, and David Lovell

Abstract We describe an investigation into how Massey University's Pollen Classifynder can accelerate the understanding of pollen and its role in nature. The Classifynder is an imaging microscopy system that can locate, image and classify slide based pollen samples. Given the laboriousness of purely manual image acquisition and identification it is vital to exploit assistive technologies like the Classifynder to enable acquisition and analysis of pollen samples. It is also vital that we understand the strengths and limitations of automated systems so that they can be used (and improved) to compliment the strengths and weaknesses of human analysts to the greatest extent possible. This article reviews some of our experiences with the Classifynder system and our exploration of alternative classifier models to enhance both accuracy and interpretability. Our experiments in the pollen analysis problem domain have been based on samples from the Australian National University's pollen reference collection (2,890 grains, 15 species) and images bundled with the Classifynder system (400 grains, 4 species). These samples have been represented using the Classifynder image feature set. We additionally work through a real world

R. Lagerstrom (✉) • Y. Arzhaeva • L. Bischof
Digital Productivity Flagship, CSIRO, Locked Bag 17, North Ryde, Sydney, NSW 1670, Australia
e-mail: ryan.lagerstrom@csiro.au; yulia.arzhaeva@csiro.au; leanne.bischof@csiro.au

K. Holt
Institute of Natural Resources, Massey University, PB 11222, Palmerston North 4442, New Zealand
e-mail: k.holt@massey.ac.nz

S. Haberle • F. Hopf
School of Culture, History and Language, H C Coombs Blg 9, The Australian National University, Canberra, ACT 0200, Australia
e-mail: simon.haberle@anu.edu.au; felicitas.hopf@anu.edu.au

D. Lovell
Digital Productivity Flagship, CSIRO, GPO Box 664, Canberra, ACT 2601, Australia
e-mail: david.lovell@csiro.au

C. Sun et al. (eds.), *Signal and Image Analysis for Biomedical and Life Sciences*, Advances in Experimental Medicine and Biology 823,
DOI 10.1007/978-3-319-10984-8_12

case study where we assess the ability of the system to determine the pollen make-up of samples of New Zealand honey. In addition to the Classifynder's native neural network classifier, we have evaluated linear discriminant, support vector machine, decision tree and random forest classifiers on these data with encouraging results. Our hope is that our findings will help enhance the performance of future releases of the Classifynder and other systems for accelerating the acquisition and analysis of pollen samples.

Keywords Pollen • Palynology • Image analysis • Statistical classification • Automation • Machine learning

12.1 Introduction

Palynologists study samples of particulates such as pollen grains to gain an understanding of the environment under which they are produced. Among other things palynology enables vegetation and climate reconstruction for the assessment of climate change and biodiversity [12, 29]. It also underpins the science in areas from allergy research to plant reproductive biology [22, 25].

Pollen grains, the microscopic bodies produced by plants as part of their reproductive cycle, exist in a wide range of morphologies. Pollen morphology is typically unique at the family level, but in some cases unique characteristics are present down to the genus or even species level. This morphological diversity makes pollen a particularly powerful tool in many branches of science, as it allows the pollen grains to be traced back to the plants that produced them. Palynology, the study of pollen grains (and spores and other microscopic biological particulates) is harnessed in many applications including:

- Reconstructing past vegetation for the purposes of investigating climate change [12], palaeoecology [29], biodiversity [25] and human activities [22].
- Allergy research and pollen forecasting for sufferers of seasonal allergic rhinitis [6].
- Determining the floral origins of honey [17, 26].
- Plant reproductive biology [16].
- Oil prospecting [15].
- Establishing plant taxonomic relationships [5].
- Forensic science [3].

However, the analysis of pollen is a slow and laborious task that involves manually preparing samples, locating and identifying pollen grains under a microscope and finally, quantifying the abundance of various species present in any sample. This is typified by an example of recent work on lake sediments from the tropical Andes, which produced a high-resolution record of vegetation and climate change spanning the past 200 ka [12]. This exceptional record showed that climate changes in the tropics over this time period were synchronous with those in the high latitude

regions of the northern and southern hemispheres. This work has thus made a major contribution to our understanding of global climate dynamics. However, to produce this impressive pollen record required the analysis of in excess of 5,000 pollen samples. Even with six highly trained analysts counting full-time, it took a whole year before the dataset was available. Since the late 1960s the palynology community has recognized the need for automation and the role it could play in accelerating the science in these areas [10, 30]. Automation would not only make obtaining the 'standard' or 'minimum' datasets required for the various fields listed above quicker and cheaper to obtain through reduced human labour hours, it is also envisaged to considerably expand the scope of pollen analysis to well beyond what is currently possible, through:

- Increasing the volume of material analysed, i.e., more pollen grains counted per sample, more samples counted, and more sites sampled (where relevant).
- Enhanced objectivity and consistency (i.e., minimising intra- and inter-analyst variations in identification – aka 'human error').
- Higher taxonomic resolution – facilitated by the potential for machines to be more sensitive to subtle morphological variation than the human eye.

There have been several efforts towards developing systems for automated pollen analysis. The majority of these employ a combination of image processing of light microscope or SEM images, with some sort of statistical or machine learning-based classifier e.g., [8, 11, 28]. Some more recent approaches have applied other technologies including flow cytometry [24], FT-IR [7], and laser scattering [18]. Nearly all of these such systems are directed towards airborne pollen monitoring associated with allergy research.

The 'Classifynder' (www.classifynder.com) developed by Massey University [14], integrates the hardware and software required to locate, image and classify slide-based pollen samples. It combines technologies from microscopy, robotics, pattern recognition, image processing and data science to form a complete, stand alone automated pollen analysis system which can be applied in any palynology lab where conventional slide-based pollen analysis is performed.

Typically, a palynologist would build up an image library of various pollen species using the Classifynder to gather images of pollen grains from a set of training or reference slides, manually classifying the images into training folders. This library is then used to train the classifier. The composition of the library will depend on the research focus of the analyst. For example, an aeropalynologist interested in seasonal changes in airborne pollen load would build up a library composed of common wind pollinated plants (i.e., pine, oak, birch, grasses, ragweed) and use this to train a classifier for use on samples of airborne pollen. In contrast, the library of a melissopalynologist interested in the floral origins of honeys would consist of pollen of common nectar-producing plants (i.e., clover, borage, thyme).

The choice of classifier is a very important issue not only in terms of accuracy, but interpretability of results. In palynology Stillman and Flenley [30] recognized a need to investigate classifier choice as early as 1996. When the existing literature

on attempts at automated palynology is examined, the highest success rates (100 % for 13 taxa) have been achieved with a neural network classifier [20], which is the justification for its use in the Classifynder system.

However, there are a number of possible shortcomings with the neural network approach. Because the native neural net strategy does not provide a measure of error for each pollen grain classification, there is no way to streamline a review of classification results. Being able to review (and correct) very obvious misclassifications (such as when a particular species is not in the library or the corruption of a grain observation) would allow for quick improvements in classification results and accelerate the phenotyping process. Therefore, despite the previous success of neural networks in pollen classification in experimental settings, there is a clear need to investigate if other classifiers are potentially more suitable for 'routine' automated classification, where unknown grains or corrupt observations are likely to be more common.

This chapter documents two comparative investigations of classifier performance on Classifynder-generated image sets. The first investigation assesses five classifiers that are typically used in modern data analytics, with regard to both accuracy of classification and interpretability of resulting classifications: neural networks [13]; linear discriminant analysis [9]; support vector machines [4]; decision trees [1]; and random forests [2]. The second investigation compares the performance of two of these classifiers (neural network and linear discriminant analysis) on sets of images of pollen from New Zealand honey samples. Honey pollen analysis is viewed as an ideal test application for an automated palynology system for a number of reasons:

- Samples are largely free from deformed or broken grains
- Minimal non-pollen debris present.
- Typically limited number of taxa present in a sample, depending on region of origin.

This is in contrast to other types of samples (e.g., fossil samples from ancient sediments) where broken or damaged pollen grains and non-pollen debris are common, and diversity can be high. Furthermore, pollen analysis of honey is becoming increasingly important in the agricultural and biosecurity sectors. Pollen analysis potentially provides a method of quality control which is particularly important for 'boutique' honeys which command premium prices [26]. Likewise, 'honey forgery' and disguising country of origin of honeys has become more prevalent, following the placement of import bans on major honey producers such as China [26].

12.2 Data

The Classifynder's digital microscopy and software system produces 43 characteristic features for each pollen grain detected in a microscopy slide. The camera scans

the slide in low resolution looking for candidate pollen grains. Candidate grains are at this point assessed as to whether they are debris or genuine pollen grains. Once a candidate is deemed to be a genuine pollen grain, a high resolution image is taken at nine different focal depths and a composite image is created. The composite image is converted into hue, lightness and saturation space. Using only the lightness values, the pollen grain is segmented from the background by an edge detector followed by filling the interior. Image feature measurements are then computed from the segmented shape and the lightness values within the shape. The image feature categories, and the number of features are: Geometry (3), Histogram (2), Moments (7), Grey Level Co-occurrence Matrix (5), Grey Gradient Co-occurrence Matrix (12), Gabor (8), and Wavelets (6). More details of these image features are available in Zhang et al. [33].

Two datasets were available to us to gain a better understanding of the classification capabilities of the Classifynder. The first was provided by Massey University and contains 400 pollen grain images from 4 different species. This data set comes bundled with the Classifynder system to help users gain an understanding of how the analysis part of the system operates. The second was provided by the Australian National University from their pollen reference collection and contains 2,980 pollen images from 11 species. The 11 species were selected as common to the Canberra region in Australia. Table 12.1 summarises the data set while Fig. 12.1 shows example images from all species.

Initial exploratory analysis was carried out to assess the correlation structure between the image feature measurements and to see if there were outlying obser-

Table 12.1 A summary of data used in this chapter. The first column is the species name. The second column is the data source, *ARC* Australian National University Reference Collection and *CTS* Classifynder Test Set. Column three is an abbreviation for species. The final column is the number of image samples (Reprinted with permission from [19]. Copyright 2013, AIP Publishing LLC)

Species name	Source	Abbreviation	#Images
Acacia ramoissima	ARC	AR	77
Atriplex raludosa	ARC	AP	341
Brachyglottis huntii	CTS	AS	100
Casuarina littoralis	ARC	CL	172
Disphyma papillatum	CTS	DI	100
Dracophyllum arboreum	CTS	DR	100
Euphorbia hirta	ARC	EH	172
Eucalyptus fasciculosa	ARC	EF	192
Isoetes fusilla	ARC	IP	715
Myrsine chathamica	CTS	MY	100
Nothofagus cunninghamii	ARC	NC	113
Nothofagus discoidea EV	ARC	NE	172
Nothofagus ciscoidea PV	ARC	NP	504
Olearia algida	ARC	OA	121
Phyllocladus aspleniifolius	ARC	PA	122

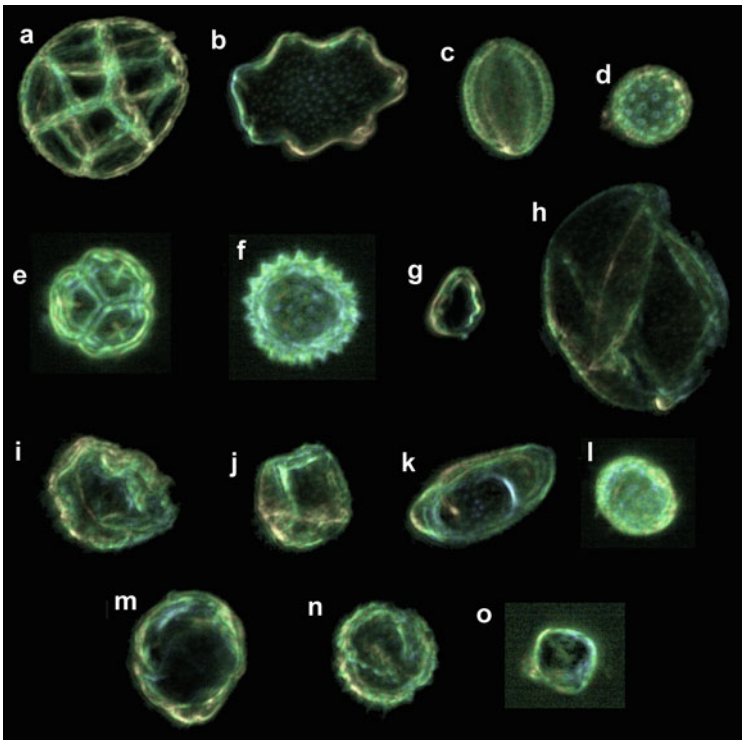


Fig. 12.1 Sample images from various species. (a) *Acacia ramoissima*, (b) *Nothofagus discoidea EV*, (c) *Euphorbia hirta*, (d) *Atriplex paludosa*, (e) *Drocophyllum arboreum*, (f) *Brachyglottis huntii*, (g) *Eucalyptus fasciculosa*, (h) *Nothofagus cunninghamii*, (i) *Isoetes pusilla*, (j) *Phyllocladus aspleniifolius*, (k) *Nothofagus discoidea PV*, (l) *Disphyma papillatum*, (m) *Casuarina littoralis*, (n) *Olearia algida*, and (o) *Myrsine chathamica* (Reprinted with permission from [19]. Copyright 2013, AIP Publishing LLC)

vations amongst the data. To investigate the correlation structure of our feature set, all observations with labelled species were considered. Looking at the correlation structure and ignoring the species labels could disguise potential discriminability, so correlation between features was examined within species and the minimum over species considered. Using this conditional correlation type approach, it was found that within species correlation differed from overall correlation for one species only, *Myrsine chathamica*. The conditional correlation between seven grey gradient co-occurrence matrix (GGCM) texture measures was 0.98 or above, while the two first level wavelet features was also 0.98. With the data sets we have, removing five of the GGCM features and one of the first level wavelet features may lead to more simple classification models with improved parameterisations. We assessed classifier performance with and without removing correlated features.

The influence of outliers on classification depends on the classifier. In an attempt to identify potential outliers in our training sets, we first scaled each variable

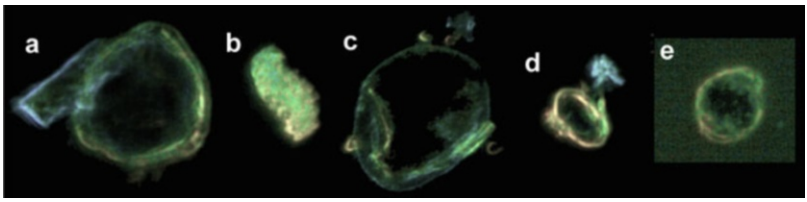


Fig. 12.2 Five sample outlying images from different image features. (a) Geometry, (b) Histogram, (c) Moments, (d) Grey Level Co-occurrence Matrix, and (e) Gabor (Reprinted with permission from [19]. Copyright 2013, AIP Publishing LLC)

to have zero mean and unit standard deviation. We found 20 observations with absolute value greater than 10 standard deviations from zero. A selection of five outlying images is shown in Fig. 12.2. They are based on geometry (NC), histogram (IP), moments (EF), GLCM (NC) and Gabor (MY). Because we found just 20 observations from 3,290 that could be characterised as outliers we take the approach of not discarding outliers as the impact on classifier accuracy would be minor.

12.3 Classification Models

Here we assess five classification models on the labelled species images: neural networks (NN), linear discriminant analysis (LDA), support vector machines (SVM), decision trees (DT) and random forests (RF). This selection of models was chosen to span linear, non-linear and tree based classifiers and represents a typical set of tools a data analyst might use to investigate classification problems in, for example, the data mining area. The analyses were performed in the statistical programming language R [27] where the classifiers are available in the nnet, MASS, kernlab, rpart and randomForest packages.

Our motive for comparing the classifiers is to ascertain whether there is a particular classifier that is especially well suited to pollen data compared to others. The developers of the system have indicated that one of the shortcomings of the neural net is the inability to measure the strength of individual classifications for the purpose of assisted reviewing. With this in mind, a question of particular relevance is how a simple linear classifier (where per observation diagnostics are available) compares to a 'black box' classifier like a neural net.

Our strategy for assessing classifier performance begins by choosing one of the three data sets: the Classifynder test set (CTS), the ANU reference collection (ARC) or the combination of both (COMB). From there we consider a data set where all feature measurements were included (FF) and also when correlated feature measurements were removed (LF). Once the data set was determined, all feature measurements were scaled to have zero mean and unit variance. The data set was then randomly split into equally sized training and test sets. Then each of the five

classifier models was built using the training data. Confusion matrices were formed based on the test set and an error measure computed. A performance measure was defined as the number of correct classifications divided by the total number of image observations. The test data was then used for training and the training data for testing in a two-fold cross validation. This process was repeated ten times and the average performance measure and confusion matrix calculated.

The neural network used was the feed-forward with single hidden layer network. The number of units in the hidden layer was set to 3, initial random weights set to 0.1 with decay 0.0005 and the maximum number of iterations equal to 600. The linear discriminant analysis model used all of the input features (i.e., we did not attempt dimension reduction via principal components or other means). The support vector machine used the C classification model. A Gaussian radial basis kernel was employed with a sigma equal to 0.1 while the cost of constraints violation parameter, C, was set to 10. For the decision tree model, no surrogates were used in the splitting process. For the random forest model, the number of trees parameter was set to 500 while the number of variables randomly sampled as candidates at each split was set to 3.

12.4 Model Performance

Model performance for the five classifier models on the six data sets is summarized in Table 12.2. The most obvious issue at first glance is the poor relative performance of the DT model on the ARC and COMB data sets. The DT model's performance is comparable to the other models on the CTS data. This may indicate the DT model is not suitable for classification when the number of species is larger. Of the other models, the LDA, SVM and NN models have the best performance over the six data sets. The SVM model outperforms the others on the both forms of the ARC data while the LDA approach outperforms the others on the both forms of the

Table 12.2 This table summarizes the performance of the five classification models over the six data sets. The performance measure is the sum of the diagonal elements of the corresponding confusion matrix divided by the number of observations. The underlined elements correspond to the best performance for each data set (Reprinted with permission from [19]. Copyright 2013, AIP Publishing LLC)

Data		NN	LDA	SVM	DT	RF
ARC	FF	0.80	0.82	0.83	0.74	0.81
	LF	0.79	0.83	0.84	0.74	0.82
CTS	FF	0.97	0.96	0.93	0.94	0.95
	LF	0.97	0.98	0.93	0.93	0.96
COMB	FF	0.80	0.84	0.82	0.75	0.83
	LF	0.80	0.83	0.82	0.74	0.82

Table 12.3 Full confusion matrix for the NN model on the COMB FF data. The rows correspond to ground truth while the columns represent the classifications results (Reprinted with permission from [19]. Copyright 2013, AIP Publishing LLC)

	AP	AR	AS	CL	DI	DR	EF	EH	IP	MY	NC	NE	NP	OA	PA
AP	87	0	0	2	0	0	3	4	0	0	0	0	0	3	1
AR	0	100	0	0	0	0	0	0	0	0	0	0	0	0	0
AS	1	0	79	0	2	12	0	2	0	0	0	4	0	0	0
CL	0	0	0	49	0	0	0	0	34	1	0	0	15	2	0
DI	0	0	0	0	98	0	0	0	0	2	0	0	0	0	0
DR	0	0	4	0	3	93	0	0	0	0	0	0	0	0	0
EF	3	0	0	0	0	0	80	0	11	0	0	2	0	1	3
EH	0	0	0	1	0	0	0	93	0	4	0	2	0	1	0
IP	1	0	0	2	0	0	2	1	88	0	1	2	1	2	1
MY	4	0	2	0	11	0	0	1	0	80	0	0	2	0	0
NC	0	0	0	0	0	0	0	2	4	0	86	2	4	3	0
NE	1	0	0	6	0	0	0	2	8	0	0	74	9	0	0
NP	0	0	0	2	0	0	0	0	2	0	0	2	93	0	0
OA	3	0	0	0	0	0	3	0	36	0	0	0	0	57	3
PA	7	0	0	5	0	0	7	3	63	0	0	1	0	10	4

COMB data and the reduced feature form of the CTS data. However, in terms of the performance measure with this data, the difference between models is slight, apart from the DT model.

A full confusion matrix for the NN model on the COMB data with all features is shown in Table 12.3. The values in the table are percentages with rows corresponding to the truth and columns to classification results, so the sum for a particular row should be 100. The NN model is used natively in the Classifynder system. Firstly, looking at the diagonal elements, the CL, OA and PA species are poorly classified with success rates under 60 % and PA in particular at 4 %. These 3 species are most frequently confused with the IP species which has the highest number of observations, 715, in the data. The number of observations for CL, OA and PA are 172, 121 and 122 respectively. The morphology and texture of the images from these species are the most similar amongst the species in the data set. Similar observations can be made from the confusion matrices for the SVM and RF models and so their confusion matrices are not displayed here. Table 12.4 shows the full confusion matrix for the LDA model on the COMB data. The two species with the lowest classification accuracy are PA with 69 % and IP with 72 %. The high number of IP observations accounts for most of the model's overall classification error. Another feature of the confusion matrix is that all species other than AR are confused with the PA species. Despite this, the error is more balanced between species which would appear to be a desirable result.

Table 12.4 Full confusion matrix for the LDA model on the COMB FF data. The rows correspond to ground truth while the columns represent the classifications results (Reprinted with permission from [19]. Copyright 2013, AIP Publishing LLC)

	AP	AR	AS	CL	DI	DR	EF	EH	IP	MY	NC	NE	NP	OA	PA
AP	84	0	0	0	0	0	4	3	1	0	0	1	0	2	5
AR	0	100	0	0	0	0	0	0	0	0	0	0	0	0	0
AS	0	0	92	0	0	5	0	0	0	0	0	0	0	0	3
CL	0	0	0	78	1	0	3	0	10	0	0	0	2	0	5
DI	0	0	0	0	98	0	0	0	0	0	0	0	0	0	2
DR	0	0	9	0	0	88	0	0	0	0	0	0	0	0	3
EF	1	0	0	3	0	0	83	0	4	0	0	0	0	5	5
EH	0	0	0	0	0	0	0	94	1	0	0	0	0	2	3
IP	1	0	0	7	0	0	3	0	72	2	0	4	0	3	8
MY	0	0	0	0	0	0	0	2	0	93	0	0	0	0	6
NC	0	0	0	4	0	0	0	1	0	1	81	2	12	0	1
NE	0	0	0	10	0	0	0	0	4	0	0	77	7	0	2
NP	0	0	0	3	0	0	0	0	2	0	0	2	92	0	0
OA	4	0	0	0	0	0	0	0	1	0	0	0	0	89	6
PA	2	0	0	4	3	0	0	0	11	0	0	3	2	7	69

12.5 Semi Automated Performance Enhancement

In practice, the classification results need not be the end point of an investigation. Typically a palynolgist would review and adjust the classification results. The LDA model presents a simple means for assisting the review stage. LDA works by transforming the data into an optimal space for discrimination. For each species, a mean value for each discriminant is then calculated. The model then measures the Mahalanobis distance between a sample and the set of mean discriminants for each species, with the lowest distance informing the choice of species for classification. This distance itself provides a measure of how far from the training data a particular sample is. Taking the ratio of lowest over the second lowest distance score gives a measure of how 'borderline' a classification decision is. So, for example, a ratio close to 0 would indicate a strong decision while a value close to 1 would indicate possible confusion. Table 12.5 shows the top 20 classification results for the LDA model on the COMB FF data ranked on decreasing values of this ratio. It shows that out of the 20 results only 4 are correctly classified. So in the context of reviewing the data, if a user were to sort their observations based on the ratio, they could easily and efficiently adjust decisions for the most borderline cases. The ratio is similar in spirit to the posterior probability which can be calculated for the LDA models and the RF models. For each classification, a posterior probability is assigned for each class. It is then possible to use this to rank the data in a similar fashion to

Table 12.5 Worst 20 classification results for the LDA model on the COMB FF data based on distance ratio (Reprinted with permission from [19]. Copyright 2013, AIP Publishing LLC)

Predicted species	Species	Distance	Ratio
NP	CL	3.86	0.99
DI	CL	13.84	0.99
PA	IP	3.72	0.99
NE	MY	7.43	0.99
AS	DR	6.79	0.99
OA	IP	2.60	0.99
OA	EF	4.08	0.99
NP	NE	4.99	0.99
CL	IP	3.85	0.99
EH	PA	3.44	0.99
IP	IP	5.21	0.99
IP	IP	4.21	0.99
CL	IP	2.72	0.99
PA	PA	2.48	0.99
PA	IP	2.16	0.99
EH	AP	3.85	0.99
NP	NP	3.28	0.99
IP	NP	3.88	0.99
IP	EF	3.45	0.99
PA	OA	6.58	0.99

the ranking, noting that a posterior probability close to 1 corresponds to a strong decision. This would allow the user to use the RF model to perform a similar type of assisted review.

Another, more automated approach is to simply exclude a proportion of the classification results based on the ratio. For example, after ranking the results on decreasing values of the ratio, one can exclude the worst N percent of the results. For the LDA model on the COMB FF data, if we exclude 20 % of the results based on this strategy, the overall performance of the classifier increases to 0.94. This compared to the performance 0.84, when all data is used, is a significant increase. If we excluded 50 % of the data the performance increases to 0.99. However, when one examines the confusion matrix corresponding to only 50 % of the data, the relative proportions of the species are modified in a reasonably substantial way. For example, the IP, NP and AP species account for 18, 15 and 10 % of the species present in the COMB FF data, respectively. When we exclude the worst 20 % of the data, the relative percentages are 16, 16 and 11 which is not too dissimilar. However, when we exclude 50 % of the data, the relative percentages are 5, 23 and 13 which is very different to the known abundances. So if the goal of a palynologist is to study relative abundance of species in a sample, one would need to find an appropriate percentage for exclusion which would preserve relative abundance.

12.6 Case Study: New Zealand Honey

In this section we work through a real world case study to demonstrate the use of the Classifynder and how various classification strategies employing neural networks and linear discriminant analysis can be incorporated into the workflow.

12.6.1 Data and Methods

The data source for this section is a suite of slides containing pollen extracted from 45 different New Zealand honeys. The pollen composition of these slides has already been determined through manual counting, with the results presented (as proportions) by Mildenhall and Tremain [23]. Across the 45 samples, Mildenhall and Tremain recognize 21 common pollen types, however, the individual classifications at the image level are unavailable. The slides also contain *Lycopodium* marker spores, which are added to determine pollen concentration per gram of honey [31]. A subset of slides were selected to supply images for a library. These slides were scanned on the Classifynder and the resulting images manually classified and sorted into folders to create a master set (Table 12.6). Insufficient numbers of images were obtained for 5 of the 21 common honey pollen types, preventing the generation of a library class for these types. The final library thus contained images of 17 taxa (Table 12.6 and Fig. 12.3). The image set for each taxon was divided into two, with one half to be used for training the classifiers, and the other half for testing.

Table 12.6 Number of images in the library. Divide by 2 to give the number of images used in the training set

Taxon	Abbreviation	#Images
Asteraceae	AS	274
Coprosma	CP	81
Echium	EC	40
Geniostoma	GE	94
Griselinia	GR	97
Ixerba brexiodes	IX	104
Knightia excelsa	KN	184
Leptospermum/Kunzea type	LE	314
Lotus	LO	250
Lycopodium	LY	325
Metrosiderous	ME	300
Poaceae	PO	49
Quintinia	QU	300
Salix	SA	239
Taraxacum	TA	155
Trifolium	TR	182
Weinmannia	WE	300

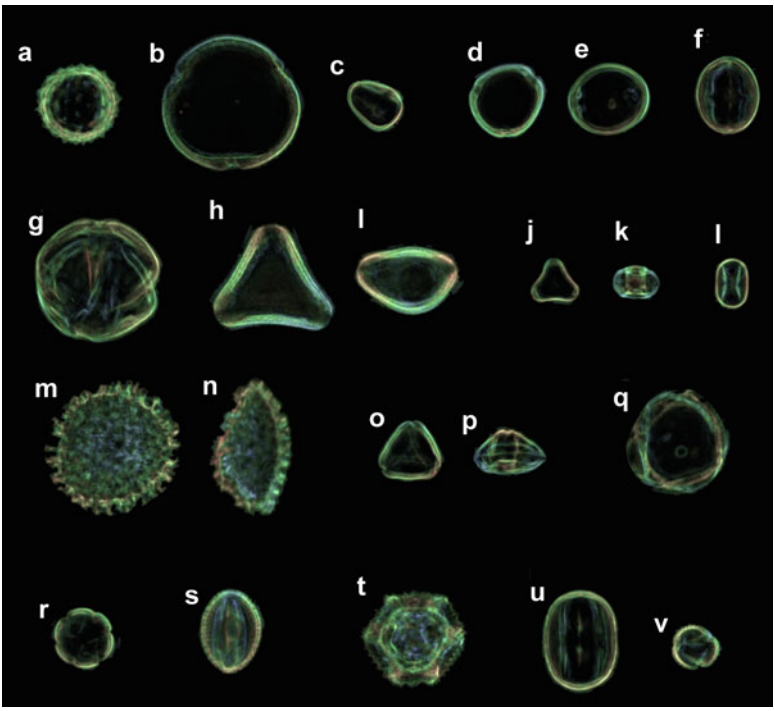


Fig. 12.3 Representative images of the 17 pollen types. Images from different angles are presented for selected pollen types. a: *Asteraceae* (AS), b: *Coprosma* (CO), c: *Echium* (EC), d, e: *Geniostoma* (GE), f: *Griselinia* (GR), g: *Ixerba brexiodes* (IX), h, I: *Knightia excelsa* (KN), j, k: *Leptospermum/Kunzea type*, l: *Lotus* (LO), m, n: *Lycopodium* (LY), O, p: *Metrosideros* (ME), q: *Poaceae* (PO), r: *Quintinia* (QU), s: *Salix* (SA). t: *Taraxacum* (TA), u: *Trifolium* (TR), v: *Weinmannia* (WE)

This training and testing set was used to train and test both the Classifynder neural network classifier and the LDA classifier.

The software of the Classifynder trains the neural network by splitting the training set provided to it and first training itself on half of this test set. It then tests its performance on the other half of the training set. It then takes any misclassified images and adds them to its training set file and retests itself in an attempt to improve performance of the neural network.

Once this automated training and testing was complete, the neural network was tested on the separate manually generated test set. Classification accuracy on the training set was 89 %. The LDA classifier was also applied to the training and test sets also achieving an accuracy of 89 %.

Once trained, both classifiers were applied to image sets gathered from three different slides from the original 45 slide set, none of which had been used as a source of images for training and testing. The three slides contain a different combination and number of the 17 taxa in the training set. These three slides

were scanned on the Classifynder. The Classifynder software sourced preliminary morphological limits from the neural network for use in identifying pollen grains on the slides at low magnification. These were then imaged using the high resolution camera. The resulting image sets were then classified by the two different classifiers (Classifynder NN and LDA).

12.6.2 Performance Assessment

In assessing the performance of the classifiers, allowances had to be made for the fact that these 'real world' datasets contained numerous images that were of objects not present in the training set. These images fell into one of several categories described in Table 12.7.

When assessing the accuracy of the classifications (i.e., proportion of images correctly classified), images assigned to these categories by the analyst were omitted from the calculations. Because both classifiers are forced, they have to classify every image as something within the training set. Therefore, because these types of image were not included in the training set, it is not reasonable (or in many cases possible) to expect correct classification. Although, in some cases images of clumps comprising grains of the same taxon were correctly classified as that taxon. Likewise, images of clumps of different taxa were sometimes classified as one of the taxa in the clump. But because at this stage there is no mechanism to adequately deal with clumps (i.e., recognizing multiple grains in the one image) we have chosen to omit these altogether.

As discussed earlier, in contrast to the NN, the LDA provides an indication of the strength of the classification of each image and these scores potentially provide a basis to further improve the results of the classification. The majority of incorrectly classified images have very high (i.e., weak) classification strength scores. When examining these, many are images which belong to the clumps, unknown, other or junk categories. One possible option to improve the classification is to automatically

Table 12.7 Descriptions of types of images obtained from the sample slide which were not included in the training set

Pollen type	Description
Other pollen	Images of rare types of pollen which were not present in the training set
Unknown pollen	Images of pollen grains in which the pollen grain was not able to be identified (due to either pollen quality pollen grain, or low-quality image)
Non-pollen	Image of an object that was not pollen
Clumps	Groups of two or more objects (usually pollen, occasionally includes junk), further subdivided into: • Clumps which contain the taxon the image was classified as • Clumps composed entirely of the taxon the image was classified as • Clumps not containing the taxon the image was classified as

Table 12.8 Classification accuracy and performance

	#Taxa	Classifynder NN	LDA ALL	LDA 80 %	LDA 50 %
Slide 2	14	80.4 %	68.4 %	73.5 %	80.8 %
Slide 12	9	76.8 %	79.6 %	88.8 %	96.7 %
Slide 26	10	74.3 %	81.2 %	86.2 %	93.4 %

Table 12.9 Comparison between proportions generated by original manual human counts (H) and manually corrected Classifynder scanning and classification (CF). The number of images were 1,587, 2,415 and 3,951 for Slide 2, 12 and 26 respectively

	AS	CP	EC	GE	GR	IX	KN	LE	LO	ME	PO	QU	SA	TA	TR	WE
Slide 2																
H	1.1	–	–	–	–	5.3	2.1	–	14.9	1.1	1.1		5.3	3.2	64.9	1.1
CF	2.1	0.1	0.0	0.2	0.0	5.1	0.4	0.6	20.6	0.5	0.5	0.1	2.9	8.3	54.9	3.7
Slide 12																
H	–	–	–	–	–	–	–	–	1.0	94.9	–	–	–	–	1.0	2.0
CF	0.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.3	95.5	0.0	0.8	0.0	0.2	0.3	2.6
Slide 26																
H	–	–	–	–	–	–	–	40.8	24.5	–	–	–	–	–	26.5	–
CF	0.2	0.1	0.0	0.0	0.0	0.1	0.0	40.3	22.8	0.0	0.0	0.0	0.0	1.1	28.4	0.1

ignore or disregard a portion of images with the highest (weakest) classification strength scores. We have assessed this approach with two subsets of the LDA classification of the initial complete image set, which take the top 80 % (LDA80) and top 50 % (LDA50) of the classified images, based on their classification strength scores, and then calculated the proportions for these in the same way as was done with the previous classifications. Results are presented in Table 12.8.

After assessing classification accuracy, any incorrectly classified images were sorted into their correct categories manually by a human palynologist, and the proportions of the different taxa as detected by the Classifynder's imaging systems calculated. These are presented in Table 12.9, and compared with the original human counts of the samples/slides, as presented in [23]. This data is a measure of how well the Classifynder can find and image the pollen using the basic morphological information extracted from the neural network, but it still has an element of manual classification through the correcting of misclassifications.

These proportions are calculated by taking the sum of all images of single pollen grains (i.e., no clump images) of a taxon (whether correctly classified or not) and dividing it by the sum of all images of single pollen grains, excluding images of non-pollen, clumps, and unknown, and also *Lycopodium* spores. Images of clumps, unknowns, other pollen and junk are also excluded, for the reasons mentioned earlier *Lycopodium* is excluded because the number of *Lycopodium* grains in each slide was not originally presented by Mildenhall and Tremain [23]. *Lycopodium* spores are added to honey samples to aid in determination of pollen concentration and their abundance has no bearing on floral composition of honey.

For slides 12 and 26, the proportions are very similar. Such minor differences between these values and those presented in [23], including the presence of additional taxa in very small amounts (e.g., AS in slide 12 and Cop in slide 26) in the Classifynder scan can be explained by differences in the area of slide examined and number of grains counted. Human analysts complete traverses of the slide, often separated by one or more fields of view (i.e., Fig. 1, page S20 [32]), until a target number is reached, in this case 500 grains [23]. The Classifynder examines every single object within a 1 cm squared area about the centre of the slide. The total number of grains examined here is typically 3–7 times that of the human counts, technically making them more statistically robust. Therefore, it is likely that minor variation between the human and machine generated proportions will be encountered. When the 95 % confidence intervals are calculated for the human counts (following the method of [21], and assuming a total count of 500 grains), the Classifynder counts fall within these limits.

The degree of difference between the original human counts and the Classifynder counts is much greater for slide 2, with the Classifynder values lying outside the 95.

The raw proportions produced by the classifiers (i.e., any misclassifications left uncorrected) are also compared with the two sets of 'true proportions' for each slide in Table 12.10. These proportions are calculated by taking the total number of images classified as a taxon, whether correct or not, is then divided by two different totals to give two sets of proportions. The first total includes all images captured, including clumps, non-pollen, other and unknown pollen, but excludes images of *Lycopodium* spores (for reasons discussed previously). The second total excludes images of clumps, non-pollen, other pollen, unknown pollen, and *Lycopodium*. The purpose of these two different totals is to investigate what impact the presence of junk, clumps and unrecognizable pollen has on the raw results and to determine whether there are any taxa which these types of images are more likely to be classified as.

The root mean square error calculated is also presented to provide an additional indication of the success of each classification.

12.6.3 Discussion

The first point of interest from Table 12.8 is that the LDA performs slightly better than the NN on 2 out of the 3 slides. When the best 80 and 50 % sets are considered, the LDA performance is significantly better than that of the NN. However, this is not true for slide 2 where the performance of the LDA is only comparable to the NN when the best 50 % of the data is retained. It is unclear why this difference occurs. One possibility is that the number of taxa present in slide 2, based on human counts, is larger (10 for slide 2 compared with, 5 and 3 for slides 12 and 26) and has an impact on classification, although, our initial assessments of classifier performance (in Table 12.2 for example) suggest comparable performance for a relatively similar number of classes.

Table 12.10 Proportions of taxa. H and CF as per Table 12.9. N = Classifynder Neural Network, L = Linear Discriminant Analysis, L80 and L5 = LDA sets with top 80 and 50 % highest classification strength scores, respectively. * denotes proportions calculated without clumps, non-pollen, unknown, and other pollen images. RM = Root Mean Square Error. Proportion falling outside of the 95 % confidence limits are underlined

	AS	CP	EC	GE	GR	IX	KN	LE	LO	ME	PO	QU	SA	TA	TR	WE	RM
Slide 2																	
H	1.1	–	–	–	–	5.3	2.1	–	14.9	1.1	1.1	–	5.3	3.2	64.9	1.1	
CF	2.1	0.1	0.0	0.2	0.0	5.1	0.4	0.6	20.6	0.5	0.5	0.1	2.9	8.3	54.9	3.7	
N	0.7	1.3	0.4	0.9	10.4	4.9	1.7	4.5	17.1	10.9	0.5	1.8	2.4	5.7	35.9	1.0	6.3
N*	0.9	0.7	0.2	0.6	8.3	4.5	0.4	2.3	21.4	3.1	0.4	2.1	3.0	7.5	43.8	0.8	3.7
L	1.5	0.0	2.0	2.2	20.4	4.3	1.8	4.2	12.1	11.3	0.3	2.8	2.4	6.6	26.6	1.2	9.5
L*	1.5	0.0	0.8	1.8	18.3	4.7	0.9	2.4	18.3	3.4	0.3	0.9	3.1	8.3	33.7	1.4	7.1
L8	1.2	0.1	1.9	2.1	17.5	4.8	1.4	4.3	15.1	10.9	0.3	2.7	2.3	7.5	26.6	1.2	8.9
L8*	1.6	0.0	0.8	1.7	14.4	5.1	0.8	2.4	21.8	2.7	0.3	0.7	3.2	9.8	33.0	1.6	6.6
L5	1.5	0.1	1.9	1.4	12.5	6.5	1.0	2.7	24.5	7.7	0.4	2.2	3.4	11.1	21.7	1.4	9.2
L5*	1.9	0.0	1.0	1.2	8.3	6.4	0.7	1.5	32.1	1.2	0.5	0.5	4.5	14.1	24.4	1.7	8.6
Slide 12																	
H	–	–	–	–	–	–	–	–	1.0	94.9	–	1.0	–	–	1.0	2.0	
CF	0.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.3	95.5	0.0	0.8	0.0	0.2	0.3	2.6	
N	0.2	0.2	0.8	0.8	1.4	0.2	1.5	18.9	2.0	65.4	0.6	2.3	0.4	0.6	2.7	2.3	8.9
N*	0.2	0.0	0.7	0.8	1.1	0.0	0.5	19.9	2.1	69.0	0.1	2.3	0.3	0.4	0.4	2.2	8.3
L	0.5	0.2	7.8	2.6	0.8	1.1	1.3	7.2	0.4	67.6	0.6	3.5	0.5	2.1	1.8	2.0	7.6
L*	0.1	0.0	8.1	2.2	0.4	0.9	0.2	7.6	0.5	73.0	0.1	3.7	0.1	0.6	0.3	2.1	6.4
L8	0.1	0.2	4.4	1.3	0.7	1.0	0.8	4.5	0.4	79.4	0.2	2.1	0.1	1.2	1.1	2.2	4.4
L8*	0.0	0.1	4.5	1.2	0.3	0.9	0.1	4.6	0.4	82.9	0.0	2.1	0.0	0.5	0.3	2.3	3.6
L5	0.0	0.2	1.4	0.3	0.4	0.7	0.2	1.4	0.4	89.9	0.0	0.8	0.0	0.9	0.5	2.9	2.3
L5*	0.0	0.1	1.4	0.2	0.0	0.7	0.0	1.4	0.4	91.2	0.0	0.8	0.0	0.5	0.4	3.0	4.4
Slide 26																	
H	–	–	–	–	–	–	–	40.8	24.5	–	–	–	–	–	26.5	–	
CF	0.2	0.1	0.0	0.0	0.0	0.1	0.0	40.3	22.8	0.0	0.0	0.0	0.0	1.1	28.4	0.1	
N	0.1	2.0	1.4	4.9	1.5	0.4	1.3	28.3	15.1	18.4	7.2	0.4	0.2	0.7	17.0	1.2	6.6
N*	0.1	1.2	1.6	2.7	1.7	0.1	0.3	29.9	18.2	14.5	7.3	0.4	0.0	0.8	19.9	1.2	5.2
L	0.3	0.6	5.5	4.5	3.6	0.9	1.1	35.4	14.5	3.7	6.0	0.4	0.3	1.9	20.2	1.0	3.9
L*	0.1	0.2	5.3	1.4	3.8	0.6	0.3	38.0	17.6	1.2	6.5	0.3	0.0	1.1	22.6	0.9	2.7
L8	0.1	0.6	3.2	4.6	2.8	0.7	0.7	36.8	15.9	3.4	6.5	0.1	0.2	1.2	22.3	0.7	3.1
L8*	0.0	0.2	3.5	1.2	2.8	0.5	0.2	39.4	18.4	0.8	6.9	0.1	0.0	1.1	24.1	0.7	2.0
L5	0.1	0.3	1.0	4.9	1.9	0.5	0.2	37.0	17.6	2.3	8.4	0.0	0.2	1.5	23.9	0.3	2.4
L5*	0.0	0.0	1.1	1.0	1.7	0.3	0.0	40.3	19.6	0.0	9.0	0.0	0.0	1.5	25.1	0.2	1.4

Another notable issue is highlighted in Table 12.10. When proportions are compared with the manual classifications, the majority are outside the 95 % confidence limits (for the human counts). Further to this, there are many instances where images have been classified as taxa which are not actually present in the

sample (i.e., 'false positives'). In some cases these false positives comprise in excess of 20 % of the total (i.e., 'LDA' classification for slide 2 in Table 12.10). False positives are particularly problematic to the ideal goal of applying automated classification to routine/commercial honey pollen analysis. Obviously, it is far from satisfactory to have the proportions of the taxa present in the sample incorrect. But reporting the presence of species in a sample which aren't actually there (or completely omitting types which are present) adds an additional layer of complication. Often, certain pollen types are specific to a particular geographic region, and presence/absence information (regardless of abundance) can be used to trace where a honey sample has come from. Therefore false positives/negatives can further influence the interpretation of the results, beyond just determining the nectar contributions. Even though reducing the images sets based on classification strength scores improves classification accuracy overall, it does not deal to the problem of false positives, with 3.7 % FP remaining in slide 26 (L5*).

In all cases, ignoring the images of clumps, unknowns, other and junk has resulted in proportions closer to actual (i.e., lower RMSE), indicating that these are in part responsible for inaccurate results. However, when the proportions for the classifications with reduced images calculated against the sum excluding the clumps, unknowns, other and junk are examined, the proportions are still outside the 95 % confidence limits in most cases. This indicates that not all error can be explained by the presence of these types of images, and that some of the error is still resulting from confusion. Some of the most common confusions in both the NN and LDA classifications (excluding clumps, junk, etc.) are between the following:

- 'LE', 'EC', 'LO' and 'ME' (e.g. slides 12 and 26)
- 'TR', 'GE', 'GR' and 'ME' (e.g. slides 2 and 26)

Confusions of images as EC, GE or GR are responsible for many of the false positive instances discussed earlier. Confusions between LE, EC, and LO seem logical, due to the similarities in size (Fig. 12.1). Likewise, confusions between LE and ME are likely related to overall morphological similarity, even though there is a reasonable size difference. Also, in the case of slide 12, ME was by far the dominant pollen type, so therefore it is likely to be best represented in misclassifications. TR and GR pollen types are also morphologically similar (Fig. 12.3), although less so with GE. Many of these more common confusables have relatively low numbers of images in their training sets (Table 12.6). Perhaps notably, EC, GE and GR, none of which are actually present in any of the samples all have less than 50 images in their training sets. Likewise, TR, which is a dominant pollen type in 2 out of the 3 slides has only 92 images in its training set. Therefore, confusions between TR, GR and GE may be explained by undertraining. LE, LO and ME have higher numbers in their training sets (greater than 120) so undertraining is likely not the problem for confusions between these three, but can explain confusions between them and EC.

Conclusions
We investigated the classification possibilities of data generated by the Classifynder, an automated imaging system for analysing pollen which can locate, image and classify slide based pollen samples. Given pollen's importance, abundance and diversity in nature, it is vital that automated systems for pollen analysis are developed and used in order to overcome the burdens of a historically manually intensive process. We looked at linear models, non linear classification and tree based classifiers from a performance and interpretability point of view. Our findings suggest that in terms of performance, the various models achieved reasonably similar results. However, we also discussed how a conceptually simple classifier like linear discriminant analysis can be exploited to review classification results in a semi-automated or automated manner. By ordering the classification results based on a metric describing how borderline a classification result is, users can efficiently delete or adjust results where classification is questionable. We also outlined an approach to automating this process by sub-setting the results based on this ordering. The benefits of taking this approach not only allow palynologists to increase their accuracy and confidence in their findings, but also accelerate the pollen phenomics process.

References

1. L. Breiman, J.H. Friedman, R.A. Olschen, C.J. Stone, *Classification and Regression Trees* (Chapman Hall, New York, 1999)
2. L. Breiman, Random forests. Mach. Learn. **45**(1), 5–32 (2001)
3. V.M. Bryant, D.C. Mildenhall, Forensic palynology in the United States of America. Palynology **14**, 193–208 (1990)
4. C. Cortes, V. Vapnik, Support-vector networks. Mach. Learn. **20**(3), 273–297 (1995)
5. C. Chun, E.A. Hendriks, R.P.W. Duin, J.H.C. Reiber, P.S. Hiemstra, L.A. de Weger, B. Stoel, Feasibility study on automated recognition of allergenic pollen: grass, birch and mugwort. Aerobiologia. **22**, 275–284 (2006)
6. A. Cristofori, F. Cristofolini, E. Gottardini, Twenty years of aerobiological monitoring in Trentino (Italy): assessment and evaluation of airborne pollen variability. Prog. Phys. Geogr. Aerobiol. **26**, 553–261 (2010)
7. R. Dell'Anna, P. Lazzeri, M. Frisanco, F. Monti, F.M. Campeggi, E. Gottardini, M. Bersani, Pollen discrimination and classification by Fourier transform infrared (FT-IR) microspectroscopy and machine learning. Anal. Bioanal. Chem. **394**, 1443–1452 (2009)
8. M.P. De Sa-Otero, A.P. Gonzalez, M. Rodriguez-Damian, E. Cernadas, Computer-aided identification of allergenic species of Urticaceae pollen. Grana **43**, 224–230 (2004)
9. R.A. Fisher, The use of multiple measurements in taxonomic problems. Ann. Eugen. **7**(2), 179–188 (1936)
10. J.R. Flenley, The problem of pollen recognition, in *Problems of Picture Interpretation*, ed. by M.B. Clowes, J.P. Penny (CSIRO, Canberra, 1968), pp. 141–145

11. I. France, A.W.G. Duller, G.A.T. Duller, H.F. Lamb, A new approach to automated pollen analysis. Quat. Sci. Rev. **19**, 537–546 (2000)
12. M.H.M. Groot, R.G. Bogota, L.J. Lourens, H. Hooghiemstra, M. Vriend, J.C. Berrio, E. Tuenter, J. Van der Plicht, B. Van Geel, M. Ziegler, S.L. Weber, A. Betancourt, L. Contreras, S. Gaviria, C. Giraldo, N. Gonzalez, J.H.F. Jansen, M. Konert, D. Ortega, O. Rangel, G. Sarmiento, J. Vandenberghe, T. Van der Hammen, M. Van der Linden, W. Westerhoff, Ultra-high resolution pollen record from the northern Andes reveals rapid shifts in montane climates within the last two glacial cycles. Clim. Past **7**, 299–316 (2011)
13. J.A. Hertz, A. Krogh, R.G. Palmer, *An Introduction to the Theory of Neural Computing* (Addison-Wesley, Reading, 1991)
14. K. Holt, G. Allen, R. Hodgson, S. Marsland, J. Flenley, Progress towards an automated trainable pollen location and classifier system for use in the palynology laboratory. Rev. Palaeobot. Palynol. **167**, 175–183 (2011)
15. C.A. Hopping, Palynology and the oil industry. Rev. Palaeobot. Palynol. **2**, 372–399 (1967)
16. S. Hu, D.L. Dilcher, D.M. Jarzen, D.W. Taylor, Early steps of angiosperm-pollinator coevolution. Proc. Natl. Acad. Sci. **105**, 240–245 (2007)
17. G.D. Jones, V.M. Bryant, Melissopalynology in the United States: a review and critique. Palynology **16**, 63–71 (1992)
18. S. Kawashima, B. Clot, T. Fujita, Y. Takahashi, K. Nakamura, An algorithm and a device for counting airborne pollen automatically using laser optics. Atmos. Environ. **41**, 7987–7993 (2007)
19. R. Lagerstrom, Y. Arzhaeva, L. Bischof, S. Haberle, F. Hopf, D. Lovell, A comparison of classification algorithms within the classifynder pollen imaging system, in *International Symposium on Computational Models for Life Sciences*, Sydney, Nov 2013, vol. 1559, pp. 250–259
20. P. Li, W.J. Treloar, J.R. Flenley, L. Empson, Towards automation of palynology 2: the use of texture measures and neural network analysis for automated identification of optical images of pollen grains. J. Quat. Sci. **19**, 755–762 (2004)
21. L.J. Maher, Nomograms for computing 0.95 confidence limits of pollen data. Rev. Palaeobot. Palynol. **13**, 85–93 (1971)
22. M. McGlone, J. Wilmshurst, Dating initial Maori environmental impact in New Zealand. Quat. Int. **59**, 5–16 (1999)
23. D.C. Mildenhall, R. Tremain, Pollen analysis of New Zealand honey. Institute of Geological and Nuclear Sciences science report, 2009, pp. 6–19
24. K. Mitsumoto, K. Yabusaki, K. Kobayashi, H. Aoyagi, Development of a novel real-time pollen-sorting counter using species-specific pollen autofluoresence. Aerobiologia **26**, 99–111 (2010)
25. B.V. Odgaard, Fossil pollen as a record of past biodiversity. J. Biogeogr. **26**, 5–16 (1999)
26. S. Petersen, V.M. Bryant, The study of pollen and its role in the honey market. Am. Bee J. **151**, 591–594 (2011)
27. R Development Core Team, *R: A Language and Environment for Statistical Computing* (R Foundation for Statistical Computing, Vienna, 2010)
28. O. Ronneberger, E. Schultz, H. Burkhardt, Automated pollen recognition using 3D volume images from fluourescence microscopy. Aerobiologia **18**, 107–115 (2002)
29. H. Seppa, K.D. Bennett, Quaternary pollen analysis: recent progress in palaeoecology and palaeoclimatology. Prog. Phys. Geogr. **27**, 548–579 (2003)
30. E.C. Stillman, J.R. Flenley, The needs and prospects for automation in palynology. Quat. Sci. Rev. **15**, 1–5 (1996)
31. J. Stockmarr, Tablets with spores used in absolute pollen analysis. Pollen et Spores **13**, 615–621 (1971)
32. W. Von der Ohe, L. Persano Oddo, M.L. Piana, M. Morlot, P. Martin, Harmonized methods of melissopalynology. Apidologie **35**, S18–S25 (2004)
33. Y. Zhang, D.W. Fountain, R.M. Hodgson, J.R. Flenley, S. Gunetileke, Towards automation of palynology 3: pollen pattern recognition using Gabor transforms and digital moments. J. Quat. Sci. **19**, 763–768 (2004)

Chapter 13
Digital Image Processing and Analysis for Activated Sludge Wastewater Treatment

Muhammad Burhan Khan, Xue Yong Lee, Humaira Nisar, Choon Aun Ng, Kim Ho Yeap, and Aamir Saeed Malik

Abstract Activated sludge system is generally used in wastewater treatment plants for processing domestic influent. Conventionally the activated sludge wastewater treatment is monitored by measuring physico-chemical parameters like total suspended solids (TSSol), sludge volume index (SVI) and chemical oxygen demand (COD) etc. For the measurement, tests are conducted in the laboratory, which take many hours to give the final measurement. Digital image processing and analysis offers a better alternative not only to monitor and characterize the current state of activated sludge but also to predict the future state. The characterization by image processing and analysis is done by correlating the time evolution of parameters extracted by image analysis of floc and filaments with the physico-chemical parameters. This chapter briefly reviews the activated sludge wastewater treatment; and, procedures of image acquisition, preprocessing, segmentation and analysis in the specific context of activated sludge wastewater treatment. In the latter part additional procedures like z-stacking, image stitching are introduced for wastewater image preprocessing, which are not previously used in the context of activated sludge. Different preprocessing and segmentation techniques are proposed, along with the survey of imaging procedures reported in the literature. Finally the image analysis based morphological parameters and correlation of the parameters with regard to monitoring and prediction of activated sludge are discussed. Hence it is observed that image analysis can play a very useful role in the monitoring of activated sludge wastewater treatment plants.

M.B. Khan • X.Y. Lee • H. Nisar (✉) • K.H. Yeap
Faculty of Engineering and Green Technology, Department of Electronic Engineering, Universiti Tunku Abdul Rahman, Kampar, Perak, Malaysia
e-mail: humaira@utar.edu.my

C.A. Ng
Faculty of Engineering and Green Technology, Department of Environmental Engineering, Universiti Tunku Abdul Rahman, Kampar, Perak, Malaysia

A.S. Malik
Department of Electrical and Electronic Engineering, Centre for Intelligent Signal and Imaging Research, Universiti Teknologi PETRONAS, Tronoh, Perak, Malaysia

C. Sun et al. (eds.), *Signal and Image Analysis for Biomedical and Life Sciences*, Advances in Experimental Medicine and Biology 823,
DOI 10.1007/978-3-319-10984-8_13

Keywords Wastewater treatment • Activated sludge • Abnormal conditions • Image segmentation • Morphological parameters

13.1 Introduction

Image processing and analysis techniques have been widely employed to characterize and monitor the activated sludge system of wastewater treatment plant in the last decade. Initially image analysis was used to find correlation between parameters extracted and physical parameters (such as sludge volume index and mixed liquor suspended solid) of the system in filamentous bulking [1]. Later the parameters were used to characterize states (normal and abnormal) by extending the correlation for some of other possible states [26]. In the literature, more attention has been given to image analysis to extract parameters as compared to image processing techniques to enhance the image, and add precision and accuracy to the image analysis parameters. In this chapter, a brief review of image acquisition systems, image processing and analysis techniques used for activated sludge system has been presented. The main objective is to explain and assess the effectiveness of image processing and analysis techniques in characterizing and monitoring of activated sludge system. In this chapter, bright field microscopy has been considered, although the techniques and algorithms may be modified and used for other types of microscopic examinations accordingly.

The image processing involves image acquisition, image processing and extraction of qualitative and quantitative parameters using image analysis techniques without human involvement. The evolution of computation capability of computers and imaging system has facilitated the use of image processing in many applications other than wastewater treatment. These include analysis of MRI (magnetic resonance imaging) [7], X-ray tomography, ultrasound imaging [33], geographical images [12], study of algal cells and muscle fibers [23, 24].

The chapter contains the following main sections:

- Introduction to activated sludge (AS) systems in wastewater treatment plants (WWTP) and its characterization of different states
- Image acquisition, preprocessing, segmentation and analysis for AS WWTP

13.2 Introduction to as Systems in WWTPs

As awareness about the role of micro-organisms in different type of diseases and toxic substances increased, the need of wastewater treatment plants arose, which are now quite wide spread. With the start of wastewater treatment practice at the beginning of twentieth century [4], 21,594 publicly owned treatment works (POTWs) have been providing the services of wastewater collection, treatment, and disposal to 226.4 million people in US by 2008 [6]. Wastewater coming

Table 13.1 Sources of toxic waste [4]

Industry	Toxins
Coal processing	Phenolic compounds, ammonia, cyanide
Petrochemical	Oil, petrochemicals, surfactants
Pesticide, pharmaceutical, and electroplating	Toxic metals such as cadmium, copper, nickel, zinc

from domestic sewage or food industry contains non-toxic constituents, whereas the wastewater originating from industries contains toxic substances enlisted in Table 13.1. The domestic wastewater comprises of 40–60 % proteins, 25–50 % carbohydrates, 10 % fats and oils, a large number of trace organic compounds of urea, pesticides, surfactants, phenols and priority pollutants, and metals, nonmetals, benzene and chlorinated compounds [4]. The major part of organic containments is biodegradable and constituted of proteins, amino acids, peptides, carbohydrates, volatile and fatty acids and their esters [8, 28].

13.2.1 Characterization of Wastewater

There have been a different set of parameters to characterize the organic and inorganic content of wastewater for different perspectives. For example, SVI is a good indicator of settle-ability. The following are the parameters often found in literature to characterize wastewater and/or activated sludge.

13.2.1.1 Odor

Some of the organic and inorganic compounds which may cause odor are:

- Geosmin (produced by bacteria doing decomposition)
- Methyl-isobarneol (MIB) (produced by bacteria doing decomposition)
- Iron
- Manganese
- Hydrogen sulphide

General classes of odors, called the reference odors, are:

- Spicy
- Flowery
- Earthy
- Grassy
- Vegetable
- Musty
- Chemical
- Disagreeable

13.2.1.2 Turbidity

Turbidity is the murkiness, cloudiness or haziness of wastewater resulted from colloidal suspended solids, suspended solids and micro-organisms; and, it is measured by light scattered by the suspended particles, in nephelometric turbidity unit (NT).

13.2.1.3 PH

PH is a measure of acidity, caused by CO_2, metals salts, chlorine and acids; and, is used to assess coagulation, stabilization and softening of wastewater.

13.2.1.4 Dissolved Oxygen

Dissolved oxygen (DO) is the oxygen dissolved in wastewater, resulted from:

- Diffusion from air
- Aeration of wastewater
- Product of photosynthesis

The oxygen is used by micro-organisms for decomposition and needed for some organisms to stay alive inside and outside the flocs. It is measured by DO meter fitted with oxygen sensing probes.

13.2.1.5 Biochemical Oxygen Demand

Biological Oxygen Demand (BOM) is a measure of oxygen required by micro-organisms to decompose organic (carbonaceous BOM) and inorganic (nitrogenous BOM) matter in wastewater [4]. High BOM implies more content of matter to be decomposed, more growth of microbes and consequently, low DO.

13.2.1.6 Chemical Oxygen Demand

Chemical Oxygen Demand (COD) is a measure of oxygen required to oxidize organic carbon in wastewater completely. High COD implies more content of organic matter to be oxidized. If COD is much greater than BOM, it means the sample contains large amount of not-easily-biodegradable organic matter [4].

13.2.1.7 Total Organic Carbon

Total organic carbon (TOC) is the concentration of carbon atoms bound within organic content of wastewater. TOC is used to detect organic content in wastewater

plant effluent to assess quality of water for public health. It is measured by oxidation of the sample, and determining the CO_2 released by non-dispersive infrared detector [4].

13.2.1.8 Total Solids

Total solids (TVS) is the sum of organic and inorganic dissolved, settle-able and suspended solids.

13.2.1.9 Total Suspended Solids

Total suspended solids (TSSol) is a measure of organic and inorganic suspended solids including microbes.

13.2.1.10 Volatile Suspended Solids

Volatile suspended solid (VS) is the organic constituent of TSSol, which includes inorganic matter and dead or alive microbes.

13.2.1.11 Food to Micro-organisms Ratio

Food to micro-organism ratio (F:M) is the ratio of organic load and volatile suspended solid. In other words, it is proportional to the ratio of COD (or BOM) and VS. The standard procedures to measure the above parameters and others have been rigorously explained in [3].

13.2.2 Wastewater Treatment Processes

The main objectives which are to be achieved in wastewater treatment are as follows:

- Decomposition of organic and inorganic content (proteins, carbohydrates and fats) into low energy inorganic substances (water, nitrate, sulphate) by microbes
- Removal or modification of the low energy inorganic substances
- Removal of micro-organisms
- The effluent water should be least possible murky and contains no substances which are harmful to health and environment

Wastewater treatment operations are classified into three broad groups [35]:

- *Physical:* screening, sedimentation, filtration and flotation
- *Chemical:* disinfection, adsorption, and precipitation
- *Biological:* biodegradation and removal of nutrients

A wastewater treatment plant may consist of the following stages [17]:

- *Preliminary treatment:* Preliminary treatment involves removal of debris and coarse material of larger size which may clog the equipment of the wastewater treatment plant.
- *Primary treatment:* It may include the physical processes of screening, grit removal, flotation and sedimentation (primary clarifier).
- *Secondary treatment:* This stage may comprise of biological processes (such as aerobic and/or anaerobic process, trickling filter, rotating biological contactors and aeration tank), physical process of sedimentation (secondary clarifier) and chemical (such as disinfection).
- *Tertiary (advanced) treatment and disinfection:* This step is meant to remove suspended and dissolved solids left after the secondary treatment and give high quality effluent. This may include nitrogen removal (nitrification-denitrification, selective ion exchange, breakpoint chlorination, gas stripping, overland flow), phosphorous removal (chemical precipitation), suspended solids removal (chemical co-angulation, filtration), dissolved solids removal (reverse osmosis, electro dialysis, distillation) and, organic and metal removal (carbon adsorption).

Both aerobic and/or nonmelodic processes are being used in secondary wastewater treatment for domestic and industrial influent, depending upon the nature of its nature. Influent with greater amount of biodegradable material is treated with aerobic process and that with less amount of the material with anaerobic process. As the main objective here is about activated sludge aerobic WWTP, so the description of anaerobic wastewater treatment has been avoided here.

13.2.3 Aerobic Wastewater Treatment

Aerobic wastewater treatment, also termed as activated sludge process, uses a suspended growth of micro-organisms to oxidize organic and certain inorganic matter to CO_2, H_2O, NH^{4+} and new biomass. Air for the oxidation comes from diffusion or mechanical aerator. The microbes form flocs which settle in the secondary clarifier. So the activated sludge process essentially consists of Aeration tank and clarifier as shown in Fig. 13.1.

In aeration tank, effluent and return activated sludge from secondary clarifier are mixed to form mixed liquor. The recycling process increases the mean cell residence time (MART) much greater than the hydraulic retention time (HART), thereby

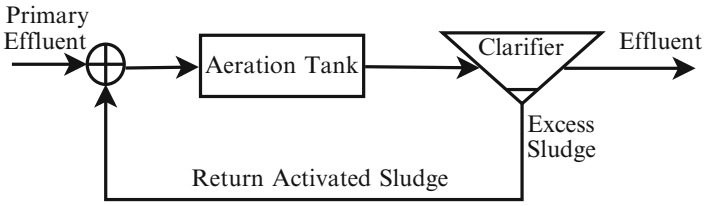


Fig. 13.1 Activated sludge system (Adapted from [4])

utilizing the microbes more effectively to oxidize organic matter in relatively short time [4]. In sedimentation tank or clarifier, the flocs formed by micro-organisms after oxidation settle to form sludge. A portion of sludge is returned to the aeration tank as return activated sludge (RAS). There are other variants of aeration such as extended or step aeration etc.; but their description has been avoided here in order to keep the description brief.

Following are some definitions specific to activated sludge process [4]:

- *Mixed Liquor Suspended Solids:* Total suspended solids (TSSol) in the aeration tank is termed as mixed liquor suspended solids (MLSS).
- *Mixed Liquor Volatile Suspended Solids:* Voltile suspended solids (VS) in aeration tank is called mixed liquor volatile suspended solids (MLVSS).
- *Food to Micro-Organism Ratio:* Food to micro-organism ratio (F:M) for activated sludge process is given by

$$F : M = \frac{Q \times BOD}{V \times MLVSS}(gBoD/gMLVSS/day) \tag{13.1}$$

  where Q is primary effluent rate in million gallons per day (MGD), and V is volume of aeration tank in gallons.
- *Hydraulic Retention Time:* Hydraulic retention time (HART) is the mean time the primary effluent stays in the aeration tank. It is given by

$$HRT = \frac{V}{Q}(day) \tag{13.2}$$

- *Sludge Age or Mean Cell Retention Time:* Sludge age or mean cell retention time (MART) is the mean residence time micro-organisms in the activated sludge system.

$$MCRT = \frac{MLSS \times V}{S_e \times Q + S_w \times Q_w}(\text{day}) \tag{13.3}$$

  where S_e is suspended solids in primary effluent, Sw is suspended solids in wasted sludge and Q_w is the rate of wasted sludge measured in (MGD).

- *Sludge Volume Index:* Sludge volume index (SVI) which is reciprocal of another parameter called sludge density index) is defined as the volume in mL occupied by 1 g of activated sludge preceded by settling of the mixed liquor for 30 min. It is determined by using the following formula:

$$SVI = \frac{SSV_{30} \times 1{,}000}{MLSS} (\text{mL/g}) \tag{13.4}$$

where SSV_{30} is volume of settled sludge in Imhoff cone containing 1 L sample of mixed liquor noted after 30 min. Activated sludge with low SVI and high sludge density index (SDI) has better settleability.

13.2.3.1 Abnormal Conditions

The following are the abnormal conditions which may occur in activated sludge process [4]:

- *Dispersed Growth:* Dispersed growth is characterized by high turbidity of primary effluent, disability of bacteria to flocculate, and dispersion of micro-organisms as small clumps; and, may be caused by high BOM, less DO, and/or toxicity.
- *Viscous or Zoogleal or Non-filamentous Bulking:* This kind of bulking is characterized by reduced settling and compaction rate; and, is resulted from excess of polysaccharides produced by bacteria.
- *Pinpoint Flocs:* The abnormal condition of pinpoint flocs or pin flocs is recognized by low sludge volume index (SVI), turbid effluent and small, compact, weak and roughly spherical flocs [4]; and, it is caused by lack of filamentous bacteria which act as backbone in forming the flocs.
- *Rising Sludge:* The abnormal condition of rising sludge or blanket rising occurs by excessive denitrification in anoxic zone of the secondary treatment where aeration tank is replaced by a sequence of anaerobic-anoxic-aerobic in order to remove nitrogen and phosphorus. The activated sludge flocs get attached to the nitrogen and float on the surface in secondary clarifier, ultimately leading to increased turbidity, high BOM and less settling ability.
- *Filamentous Bulking:* Filamentous bulking is caused by overgrowth of filamentous bacteria. Though the bacteria act as backbone in forming the flocs but their excess growth affects the settling ability and compactness of the flocs, resulted into high SVI and clear supernatant.
- *Foaming or Scum formation:* Foaming may be caused by non-biodegradable surfactants or detergents (white, frothy, billowing foam), or filamentous micro-organisms (thick, brown, stable foam). The micro-organisms which may cause the foaming are Nocardia, Microthrix Parvicella or type 1683. The foaming may also result into overflow of solid into settling tank or walkways [10].

A dichotomous key for identification of different type of filamentous bacteria has been referred frequently in literature [4,13] and the identification may lead to detection of the abnormal condition using microscopy and associated image analysis.

13.3 Image Processing and Analysis

In this section, first we shall give a critical overview of the existing work in image processing and analysis, that deals with the treatment of activated sludge wastewater treatment plants. Second, we shall discuss our preliminary work.

13.3.1 Image Acquisition

The digital image processing and analysis of activated sludge wastewater treatment plant is divided into four phases [21]: sample collection and preparation of slides, image acquisition, image processing and analysis. The samples to be analyzed by image processing and analysis are collected from aeration tank of WWTP and the slides are prepared according to the intended microscopy technique like bright field, epifluorescene, confocal laser scanning microscopy (CLSM) etc. The techniques of microscopy, which have been used for activated sludge samples, have been recently surveyed in quite detail by [26]. In Table 13.2, image acquisition procedures reported in the literature have been enumerated from selected research papers in order to give an overall picture. Bright field and phase contrast microscopy are more frequently used as compared to other techniques. Mesquita et al. used epifluorescence microscopy alongwith bright field microscopy to characterize four abnormal conditions of activated sludge system [25]; and Chu et al. employed CLSM to analyze the internal structure of floc [5]. Hence, there is no consensus that which type of microscopic image acquisition technique can give better results for the intended application of detection of abnormal conditions in WWTPs. In fact, it may depend on the intended study and abnormal condition being considered.

In the literature different sample sizes and number of images have been considered for microscopy especially for bright field and phase contrast microscopy; and in addition, most of the papers do not mention the sample size that is required for consistent and meaningful results. Heine et al. considered 20 images per slide but they did not mention the sample size and number of slides per sample [11]. Jenne et al. standardised that 50 images per sample are enough [15]. Though he established that hypothesis for the study of floc and filament counts only and did not mention the sample size for their study, the criteria of 50 images per sample was considered for morphological analysis as well by many authors [25, 26, 30].

Table 13.2 Selected work on image acquisition with plant nature

Authors	Nature of plant	Image acquisition
Jenne et al. [15]	Brewery WWTP	• Phase contrast microscopy • CCD video camera • Image acquisition covered humanly identified 100 flocs and 100 filaments
Heine et al. [11]	3 Lab-scale replica	• Bright field • CCD video camera • 10× magnification of microscope objective • Image acquisition covered 20 images per sample
Yu et al. [37]	Experimental setup	• Bright field • CCD video camera • Online digital system analysis system was built with micro-lense and a magnetic pump, for online measurements
Sikora and Smolka [32]	Not mentioned	• Bright field • CCD video camera
Chu et al. [5]	WWTP	• Phase contrast • CCD video camera • Confocal laser scanning microscope (CLSM) • Fluorescence in situ hybridization (FISH)
Jina et al. [16]	7 WWTPs	• Phase contrast • CCD video camera
Mesquita et al. [25]	Experimental setup	• Phase contrast • Epifluorescene microscopy • CCD video camera

Table 13.3 Review of sample sizes, area of coverslip and image resolution

Authors	Sample size (μL)	Size of coverslip (mm × mm)	Resolution (pixels)
Amaral and Ferreira [1]	50	24 × 24	786 × 576
Mesquita et al. [27]	25	20 × 20	1,300 × 1,030
Amaral et al. [2]	10	20 × 20	1,360 × 1,024

As far as image acquisition is concerned, generally microscope from Olympus or Carl Zeiss have been used; and, the software and procedures used are standardised by the respective commercial supplier. For example, in our work, we used ZEN Lite software with Carl Zeiss microscope. Most of the research papers did not mention the image resolution, and format although multiple options are available with the cameras and image acquisition softwares. Table 13.3 shows some selected papers with sample sizes, area of coverslip and image resolution. For the image acquisition done for the images used in the examples, 30 μL sample was used with coverslip size of 22 × 22 mm and resolution 2,560 × 1,920 pixels.

13.3.2 Image Preprocessing

The phase of image processing includes image preprocessing (noise removal and image enhancement) and image segmentation. Image segmentation is a traditional problem of image processing. A good segmentation leads to useful information extraction in image analysis phase. In the literature, floc area and filament length are considered to be two most important parameters to identify a number of abnormal conditions of activated sludge wastewater treatment plant [2]. So, different image processing techniques have been suggested to identify flocs and filaments in the subsequent portion of this chapter. The techniques are described and discussed in the specific context of activated sludge wastewater treatment plant with main interest in flocs and filaments.

13.3.2.1 Background Subtraction

Averaging of images is done to remove background noise. However, as the process is followed by z-stacking, only one of the z-stacked images is included in the averaging for background subtraction. Furthermore, the images selected to be averaged should have flocs located at different locations. The images with larger number of flocs and filaments are not expected to be improved with background correction. For example if the flocs in all the images are positioned at the center, the averaged background image will contain the floc like structure at the center as noise. Figures 13.2 and 13.3 show an example of background subtraction.

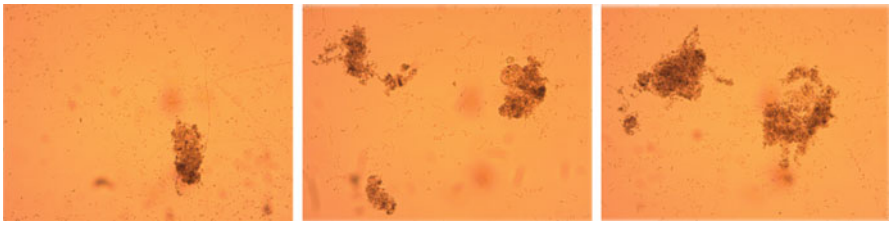


Fig. 13.2 Images before background subtraction

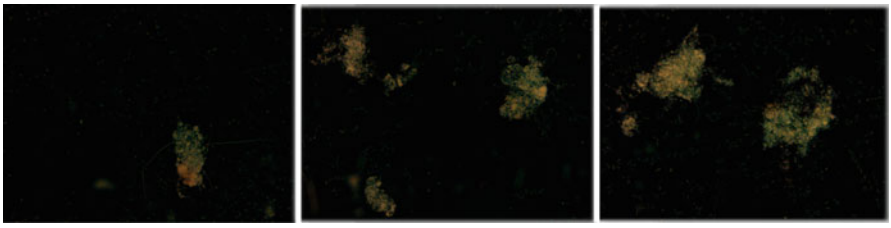


Fig. 13.3 Images after background subtraction

13.3.2.2 Z-Stacking

Z-stack is a collection of images taken at different focal planes during microscopy. It is needed because of shallow depth of field of microscope with the objective of very high magnification. The z-stack images can be fused together to reconstruct a 3-D image of the sample [9], and to make all-in-focus image by some projection method [34]. This technique has not been reported in the literature for the microscopy of activated sludge wastewater treatment.

In context of activated sludge, some filaments or portion of filaments lie in different focal planes. In order to determine precise length of the filaments, z-stacking can be used. Similar is the case with flocs. We observed in our experiments that three or four images were enough to cover all flocs and filaments. Such z-stacking can be done manually in microscopes. However, highly sophisticated microscopes with feature of z-stacking are available. For a concentrated sample, more images are expected to be needed for precise calculations. Figure 13.4 shows two background subtracted images and their z-stacked images. The z-stacking was carried out using three methods, i.e., maximum intensity, sum slices and standard deviation projection method for the sake of comparison. The comparison becomes

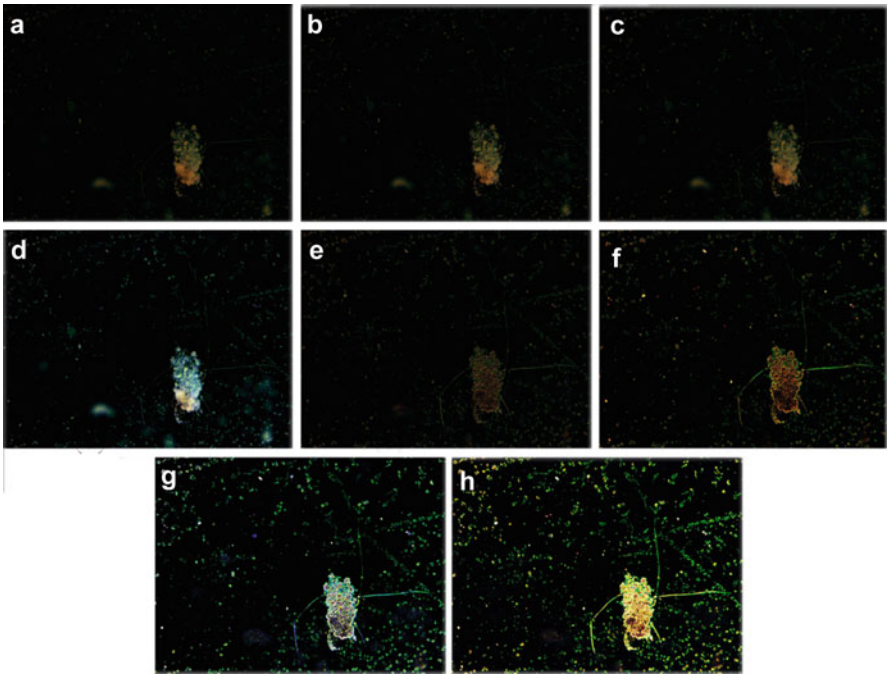


Fig. 13.4 (**a**) and (**b**) background subtracted images; z-stacking using (**c**) maximum intensity projection, (**d**) sum projection and (**e**) standard deviation projection; (**f**)–(**h**) variance filtering of (**c**)–(**e**) respectively

more explicit by using the variance filter for edge detection. Figure 13.4 obviates the fact that standard deviation projection gives the best result if variance filter is used in the later stages of image processing.

13.3.2.3 Image Stitching/Virtual Slide

Image stitching is the process of combining multiple images to form a panorama image. This method has been used for medical applications to make virtual slide and cover a larger part of an object or overall view of the sample [22, 29, 36]. This technique is also never reported in context of activated sludge wastewater treatment; although it can be quite useful in subjective as well as image analysis based quantitative identification of the state of the wastewater plant. Several algorithms have been proposed for image stitching for development of virtual slide but the discussion is beyond the scope of this chapter. We have introduced the virtual slide preparation by using the method provided as the plugin for the software ImageJ [31]. In order to explain the concept through an example, four images of a concentrated sample were acquired randomly maintaining the overlap. Then the four images were stitched through the plugin of ImageJ. The images to be stitched and the result of the stitching is shown in Fig. 13.5 where the locations of the images a to d are top left, top right, bottom left and bottom right respectively.

13.3.3 Image Segmentation

It has been observed during literature review that most of the time, the description of image processing and analysis techniques employed for activated sludge wastewater

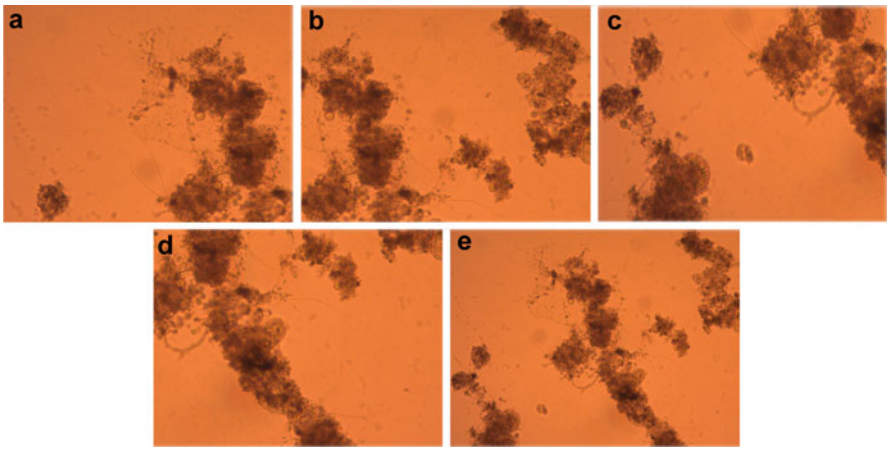


Fig. 13.5 (**a–d**) images to be stitched, (**e**) the stitched image

Table 13.4 Selected work on image segmentation for AS

Authors	Tasks done
Jenne et al. [14]	• Segmentation procedure: Histogram based thresholding (Inter-means algorithm) • Monitoring of total filament per image and mean form factor of flocs
Heine et al. [11]	• Image enhancement is done by histogram balance or median filtering • Segmentation is done by edge detection algorithm by calculation of local extrema of two dimensional intensity function of image resulting into gradient image. The procedure is followed by thresholding • Segmentation is also carried out by thresholding, followed by labeling
Sikora and Smolka [32]	• One segmentation approach is analysis of low spatial frequency component and thresholding of the smoothed image. The technique performed poorly because of irregular illumination • Separate detection steps were as follows: – Variance operator (to segment flocs) → thresholding – Laplacian operator (to segment filaments) → thresholding – Joint detection (texture discrimination operation): edge detection (Canny's algorithm) → fractal dimension → two-level thresholding
Perez et al. [30]	• Only flocs were studied • Background correction → histogram equalization → median filtering → segmentation (dilation, closing, filling, erosion) → XOR (with binary image with flocs contour) → (opening, erosion) → XOR → analysis
Lee et al. [19]	• Watershed algorithm was used for segmentation of flocs • 90–100 % quantification efficiency is achieved

treatment is very short and superfluous making the results irreproducible. Some of the research work that explains the image segmentation is tabulated in Table 13.4. We shall also discuss some different possible procedures other than those mentioned in the research papers given in Table 13.4. First of all, the previously presented example in the section of z-stacking will be extended progressing towards suggestion of a simple segmentation algorithm for flocs and filaments.

A necessary part of image segmentation is thresholding. Jenne et al. used inter-means algorithm for thresholding [14]. Many thresholding algorithms are available in literature. In the following example, we used 16 global thresholding algorithms and Li thresholding [20] was subjectively found the best for further processing.

Extending the last example, the z-stacked image from max intensity projection is segmented as shown in Fig. 13.6. First of all, Li thresholding algorithm [20] was used globally as it preserved the filaments. But one of the three filaments could not be preserved well and got lost in the debris removal step. Then intermode

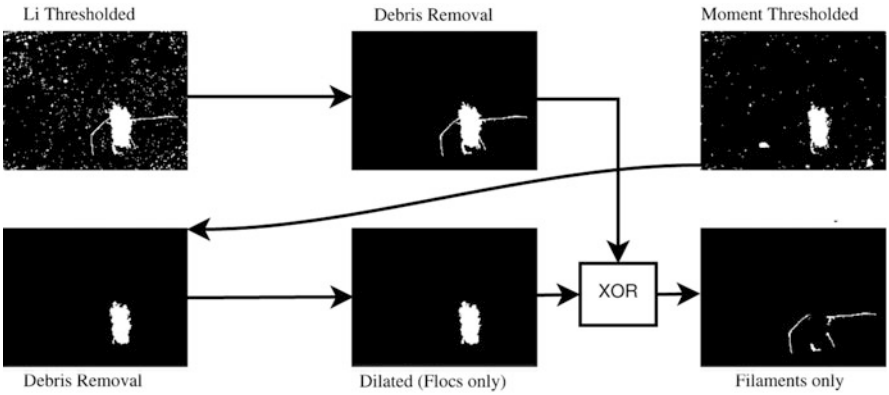


Fig. 13.6 Steps for algorithm for the example

thresholding was used, which retained the flocs only. The floc is XORed with the floc with filament to get the filaments only as shown in Fig. 13.6. The advantage of the procedure is the filaments which are attached to the flocs are segmented. The missing filament can be included by selecting some other suitable thresholding. But the procedure preserves the filaments attached to the flocs only as it considers the floc and filament the same object and considers other remotely located filaments as debris. As a conclusion, the techniques works fine only for specific images with filaments attached to the flocs. The technique can be used along with other procedures which segment remotely located filaments and flocs and miss the flocs attached to the flocs.

In order to improve the procedure described in the last example, a new algorithm was suggested with the flow diagram given in Fig. 13.7. For better segmentation, the 'floc and filament' image is obtained in one step by performing H-minima transform through the grayscale original image (hereby onwards referred to as the original image). H-minima transform eliminates regional minima in images that have a depth that is less than a given value. Continuing onwards from the last procedure, median filtering which replaces the center pixel's intensity value in the filtering window with the median value of all the points within the window is applied to the resultant image to eliminate the filament objects. Then, morphological erosion is performed to bulk up the remaining objects. Next, complementing the image produces the 'floc' image. The addition of the 'floc' image and the 'floc and filament' image produces the 'filament' image. Binary conversion is then done via a value chosen by the Otsu thresholding method. At this point, the morphological parameter of roundness is utilized to eliminate the segmented binary pixels that are not sufficiently elongated. The remaining regions construe the segmented filaments. The resulting segmentation is depicted in the images given in Fig. 13.8.

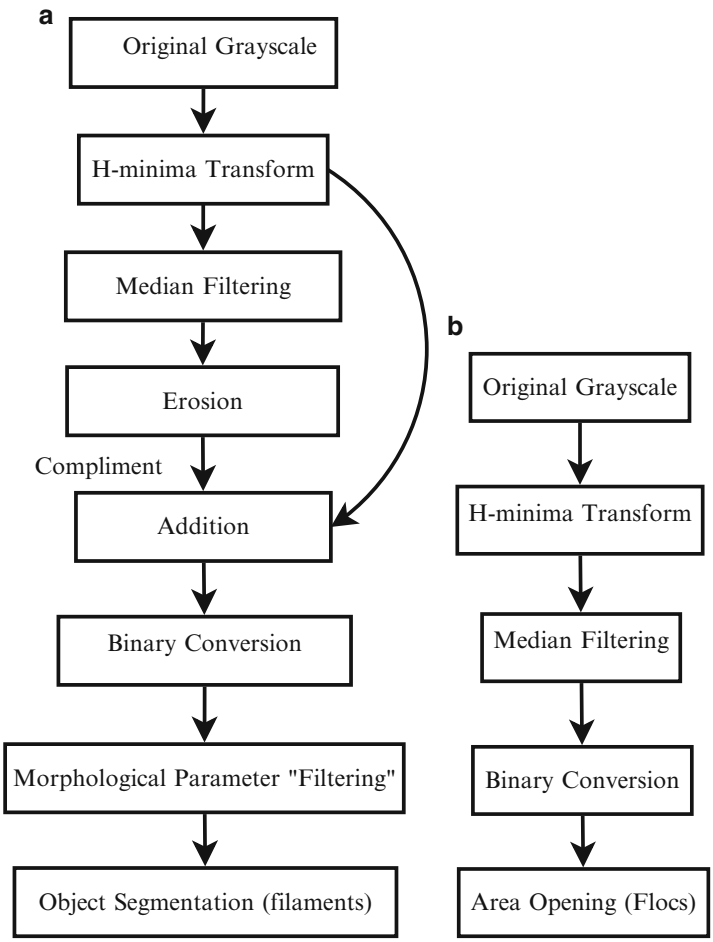


Fig. 13.7 (**a**) Filament segmentation; (**b**) Floc segmentation [18]

13.3.4 Image Analysis

In order to monitor and predict the state of activated sludge wastewater treatment plant, image analysis has been used for a number of abnormal conditions. Recently, Mesquita et al. used image analysis procedure to identify three abnormal conditions [25]. Monitoring and prediction of state of the activated sludge by image analysis is not established subjectively, but by finding the correlation with physico-chemical parameters of the plant and the morphological parameters extracted through image analysis [2, 14]. There were two types of approaches found in

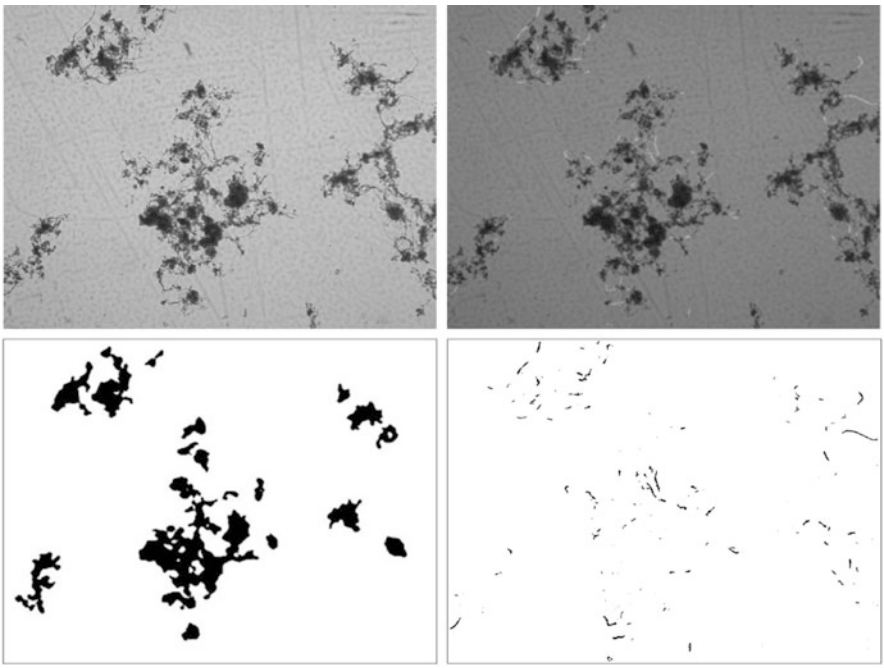


Fig. 13.8 Left to right, top to bottom, original grayscale image, superimposed filament segmentations, floc segmentation, filament segmentation

the literature about the monitoring of AS: one by finding the direct correlation between physico-chemical and image analysis parameters [1], second by "indirect" correlation by identifying coefficients of auto-regressive exogenous (ARX) model or by training the neural network [37]. As far as prediction is concerned, it is based upon the idea that properties of flocs and filaments undergo change more readily as compared to the states of the plant [11].

As shown in Table 13.5, the image analysis has been carried out broadly to differentiate between filaments and flocs, and morphological characterization of flocs. The correlation has been tried to find out between time evolution of physico-chemical and morphological parameter. The morphological parameters identified have been mentioned for the respective papers. The image analysis has been carried out in the literature without assessing the image processing algorithms. It is expected that properly assessed segmentation techniques with known probability of error can improve the correlation and help find good correlation with other morphological parameters not found suitable in the literature.

Table 13.5 Selected work on image acquisition with plant nature

Authors	Tasks done
Jenne et al. [15]	• Five simple shape descriptors are evaluated for humanly identified 100 flocs and 100 filaments • Humanly identified discriminating levels for each descriptor to differentiate floc and filaments, and determined corresponding percentage error • Radius of gyration is considered the best to differentiate between flocs and filaments • Determined required number of images of the same sample for the task of counting flocs and filaments. Mean number of flocs and filaments converge after 50 images. With 95 % confidence level, 100 images were taken
Jenne et al. [14]	• This paper extends the above paper by doing characterization of flocs and filaments along with quantification • Global features extracted: number of floc and filament and their total area • Local features: size measures and shape descriptors • Size measures (local): area, length, breadth, and perimeter • Shape descriptors (local): roundedness, form factor, reduced radius of gyration • Comparison of evolution of SVI and that of local and global features have been made • Monitoring of total filament per image and mean form factor of flocs can be used to predict filamentous bulking
Heine et al. [11]	• Correlation between sludge volume index and filament fraction • Comparison of evolution of two aeration tank with one mechanically treated: obviating the need of early detection and countermeasure in case of bulking • Evolution of fraction of microflocs with and without toxic substances • This paper partially reasons to directly relate the abnormal condition of bulking and toxics to the image analysis parameters like fraction of flocs and fraction of subflocs, without bringing in the physico-chemical parameters • Floc characterization is done by floc size, floc size distribution, structure of flocs and shape of flocs • Important survey line: “flocs have strong influence on SVI, flocculation, dewatering ability; and, they are influenced by sludge load, sludge age, the nature of substrate, for example toxic substances or the C:N:P ratio” • An important line: “characteristics of flocs respond very fast to the changes in operation conditions or operation failures”

(continued)

Table 13.5 (continued)

Authors	Tasks done
Yu et al. [37]	• Online digital system analysis system was built • Particle size distribution is monitored using laser particle size analyzer and using digital image analysis by considering equivalent diameter • Pre dominant particle size is 10–40 μm • Particle size distribution from particle analyzer and digital image analysis (DIA) are different in peaks but almost the same in shape • Find a linear relationship between fractal dimension and SS precipitation efficiency • By combining DIA result and artificial neural network (ANN), SS concentration and precipitation efficiency are predicted precisely: 16 sums of 16 different particle sizes were used as input to ANN • Identified that correlation between particle area/volume and SS is nonlinear as the density is not constant
Sikora and Smolka [32]	• Occasionally spots not belonging to filaments were found in segmented image • Filamentous bacteria were detected with very few discontinuities. Floc detection omitted some small spots inside them which can be removed by mathematical morphology
Jina et al. [16]	• The characterization of flocs is carried out using morphological, physical, chemical and extracellular polymeric substances (EPS) • Morphological parameters: floc size, filament index, fractal dimension • Different sludges had different floc sizes and morphology • Flocs with high number of filaments were also large and have relatively lower values of fractal dimensions (D_f) and high capacity of water binding
Amaral and Ferreira [1]	• There had always been filamentous bulking throughout the experiments; the relation might not valid for other modes of the plant • Morphological parameters considered: – Total aggregate number (not cut off by image boundaries) – Aggregates' mean area – Total aggregates area – Extent (area ratio of object and bounding box) – Eccentricity – Convexity (Perimeter ratio of object and convex) – Roundedness – Compactness – Solidity

(continued)

Table 13.5 (continued)

Authors	Tasks done
Amaral and Ferreira [1]	• Morphological parameters considered (Cont.): – Total aggregate number (not cut off by image boundaries) – Mean filamentous bacteria length to aggregate mean area ratio – Total filamentous bacteria length to aggregate total area ratio – Total filament length to total suspended solid ratio • All above parameters are summed up to few parameters: TL/TA, L/A, TA, solidity, eccentricity and convexity • Linear relation observed are SVI–TL/TSSol and TSSol–TA
Mesquita et al. [25]	• Three abnormal and two normal conditions were established in the experimental setup: – Filamentous Bulking – Zoogleal bulking – Pinpoint flocs – Normal condition • First, distinct patterns of SVI and MLSS are identified for the abnormal and normal conditions respectively • Correlation between ln(TL/TA) and TL/MLSS was found for four conditions • TA/Vol and TL/Vol were sketched separately for the four conditions • Damaged and viable bacteria are identified and differentiated on the basis of gram− and gram+ bacteria using fluorescence microscopy

Conclusions and Future Directions

A brief introduction of activated sludge wastewater treatment system has been presented, followed by review of existing image acquisition, preprocessing, segmentation and analysis techniques for characterization and monitoring of activated sludge flocs and filaments. Z-stacking and image stitching were suggested as a part of preprocessing. Different segmentation techniques are proposed, besides the review of techniques found in literature. For future work, optimal combination of the image processing techniques is needed for activated sludge wastewater treatment. A framework is required to characterize different image segmentation techniques used for activated sludge in terms of performance metric and capability to correlate well with the physico-chemical parameters. In future we also plan to use image processing and analysis to identify and predict abnormal conditions in activated sludge WWTPs.

Acknowledgements This work is supported by EScience Research Fund Grant funded by Ministry of Science, Technology and Innovation (MOSTI), Government of Malaysia (Project No. 06-02-11-SF0139).

References

1. A.L. Amaral, E.C. Ferreira, Activated sludge monitoring of a wastewater treatment plant using image analysis and partial least squares regression. Anal. Chim. Acta **544**(1–2), 246–253 (2005)
2. A.L. Amaral, D.P. Mesquita, E.C. Ferreira, Automatic identification of activated sludge disturbances and assessment of operational parameters. Chemosphere **91**(5), 705–710 (2013)
3. American Public Health Association, *Standard Methods for the Examination of Water and Wastewater*, 20th edn. (American Public Health Association, Washington, DC, 1998)
4. G. Bitton, *Wastewater Microbiology* (Wiley, Hoboken, 2005)
5. C.P. Chu, D.J. Lee, J.H. Tay, Bilevel thresholding of floc images. J. Colloid Interface Sci. **273**(2), 483–489 (2004)
6. EPA, Clean watersheds needs survey 2008, report to congress. Technical report, 2008
7. P.A. Fowler, C.E. Casey, G.G. Cameron, C.H. Knight, M.A. Foster, Cycling changes in composition and volume of breast during the menstrual cycle, measured by magnetic resonance imaging. Br. J. Obstet. Gynecol. **97**(7), 595–602 (1990)
8. W. Giger, P.V. Roberts, Characterization of persistent organic carbon, in *Water Pollution Microbiology*, ed. by R. Mitchell (Wiley, New York, 1978), pp. 135–175
9. A. Giusti, P. Taddei, G. Corani, L. Gambardella, C. Magli, L. Gianaroli, Artificial defocus for displaying markers in microscopy z-stacks. IEEE Trans. Vis. Comput. Graph. **17**(12), 1757–1764 (2011)
10. T. Glymph, *Wastewater Microbiology: A Handbook for Operators* (American Water Works Association, Denver, 2005)
11. W. Heine, I. Sekoulov, H. Burkhardt, L. Bergen, J. Behrendt, Early warning system for operation failures in biological stages of WWTPs by online image analysis, in *IWA Conference*, Berlin, 2001
12. G.W. Horgan, A.M. Creasey, B. Fenton, Superimposing two-dimensional gels to study genetic variation in malaria parasites. Electrophoresis **13**(11), 871–875 (1992)
13. D. Jenkins, M.G. Richard, G.T. Daigger, *Manual on the Causes and Control of Activated Sludge Bulking, Foaming, and Other Solids Separation Problems* (Water Research Commission, Pretoria, 1984)
14. R. Jenne, E.N. Banadda, N. Philips, V.J.F. Impe, Image analysis as a monitoring tool for activated sludge properties in lab-scale installations. J. Environ. Sci. Health Tox Hazard Subst Environ. Eng. **38**(10), 2009–2018 (2003)
15. R. Jenne, C. Cenens, A.H. Geeraerd, J.F. Impe, Towards on-line quantification of flocs and filaments by image analysis. Biotechnol. Lett. **24**(11), 931–935 (2002)
16. B. Jin, B.-M. Wilén, P. Lant, Impacts of marphological, physical and chemical properties of sludge flocs on dewaterability of activated sludge. Chem. Eng. J. **98**(1–2), 115–126 (2004)
17. K.D. Kerri, *Water Treatment Plant Operation (A Field Study Training Program)*, vol. 1, 6th edn. (University Enterprises, Inc./California State University, Sacramento, 2008)
18. X.Y. Lee, M.B. Khan, H. Nisar, K.H. Yeap, C.A. Ng, A.S. Malik, Morphological analysis of activated sludge flocs and filaments, in *IEEE International Instrumentation and Measurement Technology Conference*, Uruguay, 2014
19. X.Y. Lee, H. Nisar, K.H. Yeap, An approach for the segmentation and quantification of activated sludge floc blobs. Adv. Sci. Lett. **19**(5), 1372–1376 (2013)
20. C.H. Li, P.K.S. Tam, An iterative algorithm for minimum cross entropy thresholding. Pattern Recognit. Lett. **19**(8), 771–776 (1998)

21. E. Liwarska-Bizukojc, Application of image analysis techniques in activated sludge wastewater treatment processes. Biotechnol. Lett. **27**(19), 1427–1433 (2005)
22. B. Ma, T. Zimmermann, M. Rohde, S. Winkelbach, F. He, W. Lindenmaier, K.E. Dittmar, Use of autostitch for automatic stitching of microscope images. Micron **38**(5), 492–499 (2007)
23. C.A. Maltin, S.M. Hay, M.I. Delday, G.E. Lobley, P.J. Reeds, The action of the β-agonist clenbuterol on protein metabolism in invertated and denervetated phasic muscles. Biotechnology **261**(3), 965–971 (1989)
24. N.J. Martin, H.J. Fallowfield, Computer modeling of algal waste treatment systems. Water Sci. Technol. **21**(12), 1657–1660 (1989)
25. D.P. Mesquita, A.L. Amaral, E.C. Fareira, Identifying different types of bulking in an activated sludge system through quantitative image analysis. Chemoshpere **85**(4), 643–652 (2011)
26. D.P. Mesquita, A.L. Amaral, E.C. Fareira, Activated sludge characterization through microscopy: a review on quantitative image analysis and chemometric techniques. Anal. Chim. Acta **802**, 14–28 (2013)
27. D.P. Mesquita, O. Dias, R.A. Elias, A.L. Amaral, E.C. Ferreira, Dilution and magnification effects on image analysis applications in activated sludge characterization. Microsc. Microanal. **16**(5), 561–568 (2010)
28. H.A. Painter, M. Viney, Composition of domestic sewage. J. Biochem. Microbiol. Technol. **1**(2), 143–162 (1959)
29. B. Patel, T.S. Douglas, Creating a virtual slide map from sputum smear images for region-of-interest localisation in automated microscopy. Comput. Methods Programs Biomed. **108**(1), 38–52 (2012)
30. Y.G. Perez, S.G.F. Leite, M.A.Z. Coelho, Activated sludge morphology characterisation through an image analysis procedure. Braz. J. Chem. Eng. **23**(3), 319–330 (2006)
31. S. Preibisch, S. Saalfeld, P. Tomancak, Globally optimal stitching of tiled 3D microscopic image acquisitions. Bioinformatics **25**(11), 1463–1465 (2009)
32. M. Sikora, B. Smolka, Feature analysis of activated sludge based on microscopic images, in *Canadian Conference on Electrical and Computer Engineering*, Toronto, 2001
33. G. Simm, Selection for lean meat production in sheep, in *Recent Advances in Sheep and Goat Research*, ed. A. W. Speedy (CAB International, 1992), pp. 193–215
34. H. Su, F. Xing, J.D. Lee, C.A. Peterson, L. Yang, Learning based automatic detection of myonuclei in isolated single skeletal muscle fibers using multi-focus image fusion, in *2013 IEEE 10th International Symposium on Biomedical Imaging (ISBI)*, San Francisco, 2013, pp. 432–435
35. G. Tchobanoglous, F.L. Burton, *Wastewater Engineering: Treatment, Disposal and Reuse*, 4th edn. (McGraw-Hill, Boston, 2003)
36. F. Yang, Q.-H. Fan, Z.-S. Deng, A method for fast automated microscope image stitching. Micron **48**, 17–25 (2013)
37. R.-F. Yu, H.-W. Chen, W.-P. Cheng, M.-L. Chu, Simultaneously monitoring the particle size distribution, morphology and suspended solids concentration in wastewater applying digital image analysis (DIA). Environ. Monit. Assess. **148**(1–4), 19–26 (2009)

Chapter 14
A Complete System for 3D Reconstruction of Roots for Phenotypic Analysis

Pankaj Kumar, Jinhai Cai, and Stanley J. Miklavcic

Abstract Here we present a complete system for 3D reconstruction of roots grown in a transparent gel medium or washed and suspended in water. The system is capable of being fully automated as it is self calibrating. The system starts with detection of root tips in root images from an image sequence generated by a turntable motion. Root tips are detected using the statistics of Zernike moments on image patches centred on high curvature points on root boundary and Bayes classification rule. The detected root tips are tracked in the image sequence using a multi-target tracking algorithm. Conics are fitted to the root tip trajectories using a novel ellipse fitting algorithm which weighs the data points by its eccentricity. The conics projected from the circular trajectory have a complex conjugate intersection which are image of the circular points. Circular points constraint the image of the absolute conics which are directly related to the internal parameters of the camera. The pose of the camera is computed from the image of the rotation axis and the horizon. The silhouettes of the roots and camera parameters are used to reconstruction the 3D voxel model of the roots. We show the results of real 3D reconstruction of roots which are detailed and realistic for phenotypic analysis.

Keywords Root phenotyping • Root segmentation • 3D reconstruction of roots • Root tip detection • Tracking • Ellipse fitting • Zernike moments

14.1 Introduction

Root system architecture (RSA) is a fundamental component of agricultural and natural ecosystem productivity [10,16,20,38]. Several researchers [40,46] highlight the role of genes in regulating root growth rate and branching. Thus, an automated root tip detection and 3D reconstruction procedure, applied across genetic

P. Kumar (✉) • J. Cai • S.J. Miklavcic
School of Information Technology and Mathematical Sciences, Phenomics and Bioinformatics Research Centre, Australian Centre for Plant Functional Genomics, University of South Australia, Mawson Lakes, SA 5095, Australia
e-mail: pankaj.kumar@unisa.edu.au; jinhai.cai@unisa.edu.au; stan.miklavcic@unisa.edu.au

C. Sun et al. (eds.), *Signal and Image Analysis for Biomedical and Life Sciences*, Advances in Experimental Medicine and Biology 823,
DOI 10.1007/978-3-319-10984-8_14

varieties, has the potential to be a crucial plant physiological tool. There has been several approaches using image processing and 3D reconstruction for quantitative analysis of RSA [9, 37, 54]. Lobet et al. in [37] proposed a semi-automated image processing technique to streamline the quantitative analysis of growth and structure development of complex root systems. Their software, SmartRoot, is an operating system-independent freeware, based on ImageJ and relies on cross platform standards for communication with data-analysis. Clark et al. [9] developed a high-throughput phenotyping method for the tracking of 3D root traits during seedling development. Their platform has high flexibility and capacity to measure root traits at high spatial and temporal resolution. However, most of the root features are manually selected by the user and are thus at best semi-automated. In [54], Ying et al. presented a scheme for high resolution, 3D root reconstruction using the concept of visual hull.

In this chapter we present a complete system for 3D reconstruction of roots grown in transparent gel medium or roots that are grown in soil or sand, washed and suspended in water for imaging. Such a complete system, which does self calibration and is capable of automation, is not possible by just integration of already available image processing and computer vision algorithms. The feature detectors have to be especially designed and trained for this application. Thus this chapter addresses several fundamental problems of image processing and computer vision.

One fundamental problem in computer vision is that of achieving effective automatic pattern recognition, localization of desirable features and matching those features across images that are spatially and temporally separated. By solving this problem one can resolve a number of issues arising in applications to 3D reconstruction, tracking, registration, object recognition, to name a few. The definition of a desirable feature in most cases is context and application dependent [32]. A universal approach to feature detection and matching is highly unlikely to be successful. However, even if a method for widespread use could be developed, it is debatable whether it would be uniformly successful for all applications of computer vision and image processing, as the number of possible different types of desirable features is uncountable. Figure 14.1a shows an example of scale-invariant feature transform (SIFT) feature detection and matching [47] on two root images of the same root viewed from two different directions. Figure 14.1b shows an example of the proposed root tip detector and matching on the images of the same root. It is now well accepted that the context of an application has a significant influence on the choice of approach to take. With this in mind, we consider the problem of detecting root tips in images of roots of plants grown in gellan gums or suspended in a transparent medium. We propose a method of detecting root tips in 2D images based on extremal curvature scale space and the use of the statistics of Zernike moments. Our motivation for seeking to detect root tips is the essential role it plays in a program for tracking root growth development, spatially and temporally, time series images of plant roots. We chose to utilize Zernike moments (ZMs) because of their nice properties of orthogonality, rotational invariance, and demonstrated effective application to bilevel images. Teague in [45] writes that ZMs furnish an optimal means to codify the essential features of an image object that are independent of image size, centroid and relative angular orientation. Application of ZMs to bi-level

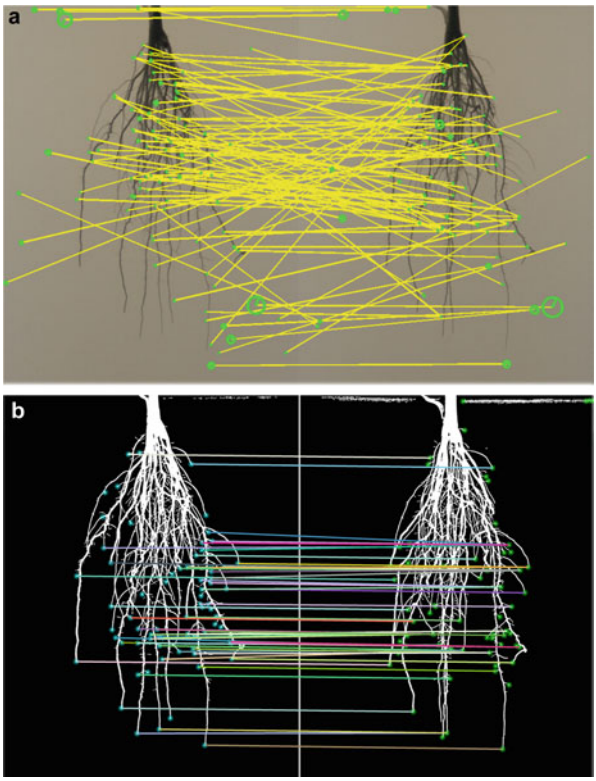


Fig. 14.1 (**a**) Example of SIFT feature detection and matching applied to root images of same root imaged from different directions. (**b**) Example of root tip detection and matching applied to root images of the same root with the proposed method

64×64 images of characters of the English alphabet has been quite successful [28]. In results we achieve as high as 100 % true detection rate at the cost of less than 2.5 % false alarm rate. On the other hand, with support vector machine we achieve 100 % true detection at the false alarm rate of 37.0 %. Detected root tips are tracked in the image sequence generated by the turntable motion. The roots are imaged by a camera rotating about an axis of rotation or the camera is stationary and the roots rotate about a fixed axis. The detections are matched across spatially separated images by co-relation of image patch around tips and nearest neighbour data association. We used a graph based algorithm similar to that proposed in [31], for tracking and track initialization. To the trajectories of root tips ellipse fitting is applied.

Ellipse fitting is another fundamental problem in computer vision and image processing. For example, ellipse or conic fitting was used to calibrate catadioptric cameras in [11] and [49], to calibrate pin hole camera for the geometry of single axis rotatory motion [22], in the segmentation of cells in microscopic images [4], segmentation of grains in [52] and in the study of galaxies in astrophysics in [41]. We use the ellipse fitting result to self calibrate the imaging sensor similar to the approach used in [23].

Previous approaches to conic fitting have all focused on minimizing a distance function to obtain the best fit to point data. An algebraic distance error was

minimized in [13] and a numerically stable version of the same was presented in [15]. A geometric distance error measure was directly minimized in [1, 2] and by maximum likelihood estimation in [8, 43]. In contrast, Kanatani proposed hyperaccuracy methods for ellipse fitting in [26] and [27], while Yu et al. in [50] proposed a new distance metric based on the intrinsic properties of ellipses. The new distance function had a clear geometric interpretation and was less computationally intensive than the geometric error. However, none of the above methods considered a non-uniform weighting of the error distance measure to account for the eccentricity of data points. Data points that are more distant from the ellipse center are considered more eccentric than data points which are closer. A mathematical definition of eccentricity is presented in Sect. 14.6. One method that considers weighting in the distance error measure calculation to make the fit more robust was adopted in [50]. Although somewhat related to the method proposed here, theirs involved a specific novel distance error measure based on the intrinsic properties of the ellipse. Our method is different from theirs as we do not propose a new error measure; instead we propose a preprocessing method of point data to which an ellipse has to be fitted. Our preprocessing is analogous to the resampling algorithm of particle filters where samples (data points) having more weight are repeated and samples with insignificant weights are dropped. After resampling of data points any ellipse fitting algorithm can be applied.

Turntable sequence of images for 3D reconstruction by self calibration is an important topic of research and development in computer vision. Here we focus on Euclidean 3D reconstruction of roots by self calibration for their phenotypic analysis. Self calibration of single axis turntable sequence has been carried out either by silhouette based approaches [14, 19, 39, 42, 51, 53] or by feature point tracking methods [6, 12, 24, 25].

Silhouette based methods are more suited for smooth, compact objects with negligible protrusions, and/or no texture on which to detect reliable point features in most of the images of the sequence. For silhouette based approach Zhang and Wong in [51] proposed a novel relation between imaged circular points $\mathbf{i}, \mathbf{j}$ and vanishing point $\mathbf{v}_x$, the axis of rotation $\mathbf{l}_s$ and horizon $\mathbf{l}_h$ and scale factor κ, the image invariants of single axis motion;

$$\mathbf{i}, \mathbf{j} \backsim \mathbf{v}_x \pm j\kappa(\mathbf{l}_s \times \mathbf{l}_h) \tag{14.1}$$

The constrain of circular points on the imaged absolute conics (ICA) $\boldsymbol{\omega}$, $\mathbf{i}\boldsymbol{\omega}\mathbf{i}^T = \mathbf{j}\boldsymbol{\omega}\mathbf{j}^T = 0$, and pole polar relationship $\mathbf{l}_s \backsim \boldsymbol{\omega}\mathbf{v}_x$, were used to compute $\boldsymbol{\omega}$ and then the internal parameters of the camera $\mathbf{K}$ was obtained by the Cholesky decomposition of $\boldsymbol{\omega}$ [17]. Mendonça et al. in [39] used the symmetry properties of the surface of revolution swept out by the rotating object to obtain the image of the rotation axis and the homography relating epipolar lines in two views. Geometric constraints on images of the rotating object are used to obtain the image of the horizon and hence the epipolar geometry of the sequence. Here the imaging set up have been designed for the biologist to get a clear view of the roots and its architecture, which some times leads to degenerate case of [39]. The set up also

avoids the specular reflection from the gellan gum container which causes problems with segmentation. Forbes et al. in [14] introduced a silhouette based method for auto calibration and 3D reconstruction of objects imaged in a twin mirror set-up. Unfortunately the geometry of root structure is such that a twin mirror set up cannot be used to image unoccluded silhouettes in all the multiple reflections. Zhang et al. in [53] related the twin mirror set-up to the circular motion and proposed methods of estimating the circular points $\mathbf{i}, \mathbf{j}$ and image of the rotation axis $\mathbf{l}_s$ of the twin mirror set up and hence auto calibration of the camera.

In [12], Fitzgibbon et al. recovered the image invariants of single axis motion and rotation angles from the fundamental matrices and trifocal tensors of adjacent views. The fundamental matrices and trifocal tensors were computed by detection and tracking of corner features on the image sequence. The initial estimates of image invariants were optimized by bundle adjustment. Jiang et al. [23, 25], fitted conics to the trajectory of the feature points and using the concept of projective geometry computed the circular points from the complex conjugate intersection of the conics. During conic fitting they applied the homography constraint, which maps the circles of the 3D space to the conics in the image plane. Jiang et al. in [24] proposed a method of computing the circular points from minimal two points and their correspondence in four images of the circular motion image sequence. A planar homography was computed from the set of points and the two eigenvectors of this homography were shown to be the images of the circular points.

Zheng et al. in [54] proposed a detailed 3D reconstruction of plant roots from turntable image sequence. They made three novel contributions: (1) Background of the root images where modelled by a harmonic function to get better segmentation of the roots, (2) Regularized visual hull which reduces the effect of jittering and refraction by ensuring consistency with one of the 2D image, and (3) Guarantee connectedness in the 3D reconstruction by minimization of a global error term. The method assumes that the internal and external parameters of the cameras are known. Camera calibration is not discussed in their paper. However our system is capable of self calibration and thus can be fully automated.

Hence forth the chapter is organized as follows: In Sect. 14.2 we briefly describe the various modules and processing stages of the system. First we briefly describe the root segmentation process in Sect. 14.3 and then in Sect. 14.4 we describe the root tip detection methodology, then we talk about root tip tracking in Sect. 14.5. In Sect. 14.6 we explain our ellipse fitting method. Then we talk about camera calibration in Sect. 14.7 and in Sect. 14.8 we show results of the different processing modules and the final results of 3D root reconstruction and finally we conclude with a brief summary and some ideas of future work.

14.2 System Design

The process pipeline and flow of data for the proposed system is shown in Fig. 14.2. Digital image of the root is acquired from a camera by imaging the root in a transparent medium on a turntable. An adaptive thresholding algorithm is applied to

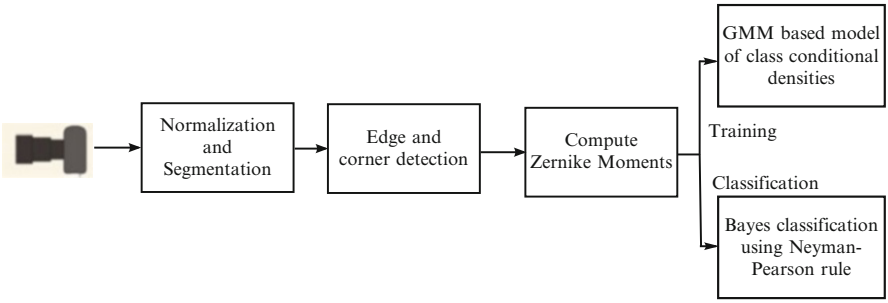


Fig. 14.2 Block diagram schematic of the proposed 3D root reconstruction system

segment the root from its background. On the bi-level image the root tip detection algorithm is applied. The detected root tips are matched across images and tracked in the image sequence generated by turntable motion. Roots which have been tracked in more that 75 % of the images of the turntable sequence are fitted with ellipses. The intersection of these ellipses is used to compute the image of the circular points and also to compute the invariants of the turntable motion, this leads to camera calibration. With camera parameters and the silhouette obtained by segmentation, visual hull reconstruction of the 3D roots are obtained.

14.3 Root Segmentation

Digital images of roots either grown in gellan gum or washed and suspended in water are converted from colour (RGB) or gray to a bi-level foreground – background image. Where foreground are the root pixels and background are the non-root pixels. There are several methods for segmentation that can be used [29, 30], however in this chapter we have used the method of adaptive thresholding. In this method an image is subdivided into smaller blocks. Within each block mean pixel intensity is computed and pixels lower than an offset to the mean are considered root pixels and rest of the other pixels are classified as background pixels (Fig. 14.3).

14.4 Root Tip Detection

Bi-level, foreground-background image $I(x, y)$ obtained above is processed for corner detection by the method proposed in [18] by He and Yung. The method is based on global and local curvature properties of the contour obtained by applying Canny edge detector to $I(x, y)$. Figure 14.4b shows the result of application of this corner detection method. It is noteworthy that the corner detection procedure results

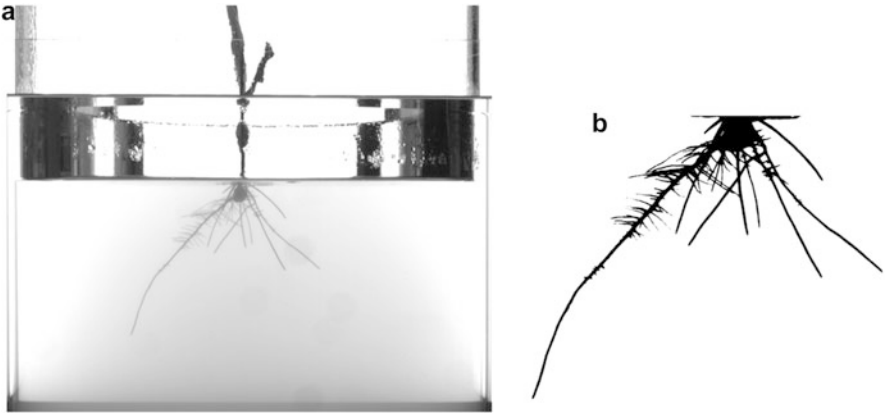


Fig. 14.3 (**a**) A *gray scale* photo of a root; (**b**) a bi-level segmented image of the same

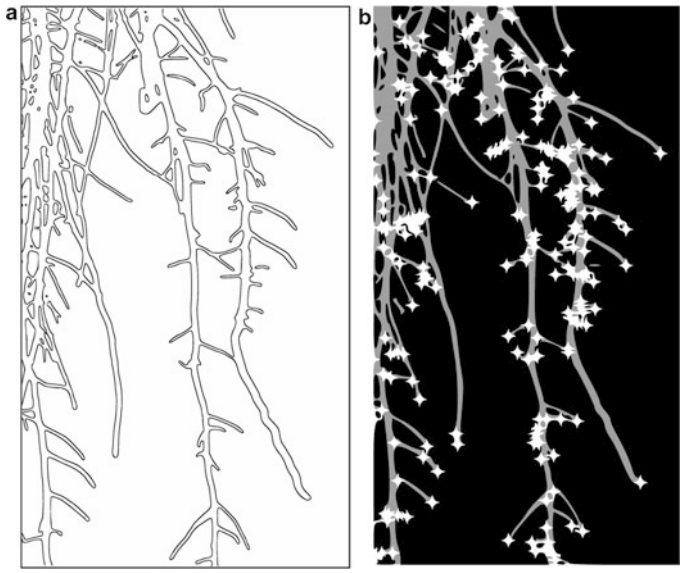


Fig. 14.4 (**a**) Edge map obtained from the bi-level image. The image has been inverted for better visibility. (**b**) Results of the corner detection method that detects all high curvature points

in the identification of corners that are either root tips, root branch points, or root cross over points. To eliminate the non-root tip corners we have used a statistical learning approach based on the statistics of Zernike moments for image patches centred around the corners. An image patch $I^c(x, y)$ of size $N \times N$ centred on detected corners are used to compute the Zernike moments both for training and classification of the patch as root-tip or non-root-tip. From a set of labelled training

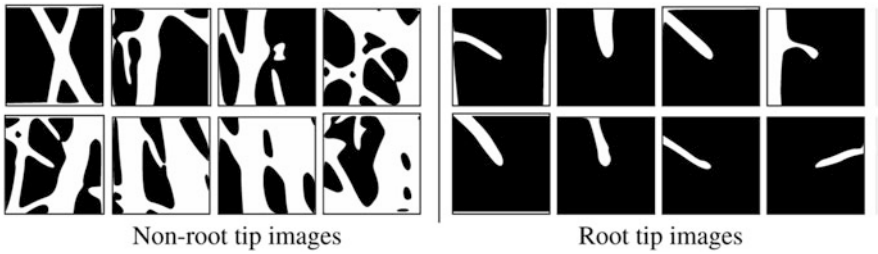


Fig. 14.5 Sample of the bi-level image patches of class non root tip and root tip obtained after preprocessing. These manually classified images patched are used to learn the class conditional densities of the feature vectors (Zernike moments) using GMM

data a sample of which is shown in Fig. 14.5, a Gaussian mixture model (GMM), for the statistics of Zernike moment for different classes are learnt. See [34] for details on Zernike moments and GMM parameter estimation. For classification after training following Bayes rule is adopted as it guarantees optimality.

From the training data the class conditional density $p(\mathbf{z}_q/rc)$ the density distribution of Zernike moment $\mathbf{z}_q$ given that the image patch is of class root tip (rc) is learnt. Similarly the class conditional density $p(\mathbf{z}_q/nrc)$ for non root tip (nrc) class is learnt.

$$p(\mathbf{z}_q/rc) = \sum_{k=1}^{K} w_k^{rc}\, g(\mathbf{z}_q; \mu_k^{rc}, \Sigma_k^{rc})$$
$$p(\mathbf{z}_q/nrc) = \sum_{k=1}^{K} w_k^{nrc}\, g(\mathbf{z}_q; \mu_k^{nrc}, \Sigma_k^{nrc}) \tag{14.2}$$

For the purpose of classifying an image patch $I^c(x, y)$ or testing the classification algorithm, the magnitudes of the nth order Zernike moments are computed. Using Bayes rule the density of class root tip given the Zernike moments is obtained as follows:

$$p(rc/\mathbf{z}_q) = \frac{p(\mathbf{z}_q/rc)\,p(rc)}{p(\mathbf{z}_q)} \tag{14.3}$$

Similarly using Bayes rule the density of class nrc given Zernike moment $\mathbf{z}_q$ is

$$p(nrc/\mathbf{z}_q) = \frac{p(\mathbf{z}_q/nrc)\,p(nrc)}{p(\mathbf{z}_q)} \tag{14.4}$$

To the above conditional density distribution we apply the Neyman-Pearson criteria [21] which leads to an optimal classification criteria. A variable η is defined which is

$$\eta = \frac{p(rc/\mathbf{z}_q)}{p(nrc/\mathbf{z}_q)} = \frac{p(\mathbf{z}_q/rc)\,p(rc)}{p(\mathbf{z}_q/nrc)\,p(nrc)} \tag{14.5}$$

For a given set of training data $p(rc) = 1 - p(nrc)$ is fixed and hence $p(rc)/p(nrc)$ is a constant and can be taken to the left hand side of Eq. (14.5) which leads to a further new variable

$$\eta^* = \eta \frac{p(nrc)}{p(rc)} = \frac{p(\mathbf{z}_q/rc)}{p(\mathbf{z}_q/nrc)} \tag{14.6}$$

If η^* is greater than an empirical threshold η_{th} then the image patch $I^c(x, y)$ and hence the corresponding corner is classified as root tip otherwise it is classified as non root tip. To generate the ROC plots the threshold, η_{th}, was varied within a sufficiently wide range. The true and false detection rates were computed for each such value of η_{th}. We thus generated ROC curves for different orders of Zernike moments and for different numbers of Gaussians in the GMM. The ROC curves for the results of classification by this method are shown in Sect. 14.8 and compared with SVM [7] classification.

14.5 Root Tip Tracking

For tracking of the root tips we have used the method of multi-target tracking described in [33]. The method cooperatively combines Kalman filter-based motion and region tracking with a nearest neighbour data association algorithm. The tracking system is fully automated requiring no manual input of any kind for initialization of tracking. The target track initialization problem is formulated as computation of shortest paths in a directed and attributed graph using Dijkstra's shortest path algorithm. This scheme correctly initializes multiple root-tip trajectories for tracking even in the presence of clutter and detection errors which occur in detection of root tips. To make the tracking problem simple and fast we have tracked only the seminal roots and not the lateral roots. Once the trajectory of the primary/seminal root tips has been obtained an ellipse fitting is applied to the trajectory points. Figure 14.6a shows an example of root tip detection and Fig. 14.6b shows the tracking of these root tips across the 72 image sequence obtained from the turntable rotation. Here we have focused on tracking the primary root tips whose detections are shown with blue asterisks in Fig. 14.6a.

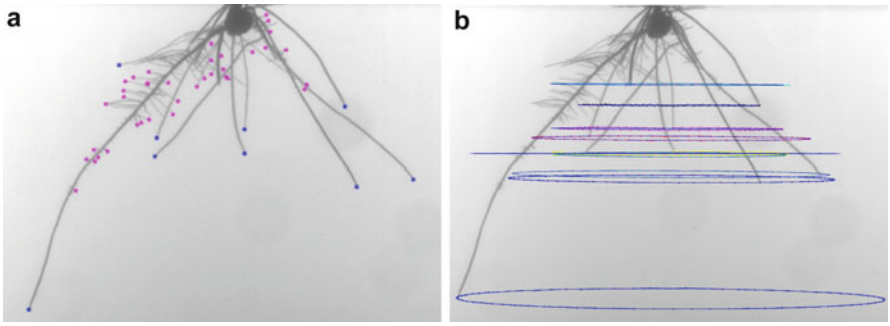


Fig. 14.6 Detection and tracking of root tips and ellipse fitting to the trajectories of the root tips are shown in these images. Due to high eccentricity of ellipses the errors in ellipse fitting are large and will lead to significant errors in estimating the image invariants of the single axis motion just by feature tracking. Therefore we propose an improved ellipse fitting method, where data points from the trajectory are pre-processed so that data points with higher eccentricity are replicated to improve the ellipse fitting results (Images courtesy of [36])

14.6 Ellipse Fitting

The eccentricity of an ellipse is defined by $\Xi = \sqrt{\frac{a^2-b^2}{a^2}}$, where a is the semi-major axis and b is the semi-minor axis of the ellipse and $0 \leq \Xi < 1$. The value $\Xi = 0$ corresponds to a circle and the value $\Xi = 1$ corresponds to a straight line. To measure the eccentricity of data point sets to which an ellipse is to be fitted, we introduce the following pointwise eccentricity variable:

$$\xi = \frac{d_a}{d_b + d_a}, \tag{14.7}$$

where d_a is the orthogonal distance of a data point to the minor axis and d_b is the orthogonal distance to the major axis. This variable, whose values range between 0 and 1, takes on larger values for data points that are more distant from the minor axis than from the major axis. In [35] an interesting observation is reported, that is data points with higher eccentricity have higher root mean square error (RMSE) when compared with data points with lower eccentricity. Thus if the contribution of data points with higher eccentricity are increased compared to those with low eccentricity, better ellipse fitting can be obtained. This is achieved by resampling data points such that data points with high eccentricity are repeated.

Resampling is a process used in particle filters to avoid the problem of particle degeneration [3]. In this process the particles having greater weight are repeated and particles with insignificant weights are dropped, with the overall number of particles being preserved. Algorithm 1 gives the pseudo code for this method. In the present application, we adopt the same concept to have multiple replicates of those data points having high point wise eccentricity values. A weight is assigned to each data

point r using the weight function $w^r = e^{\xi^r}$, where $\xi^r = \left(\frac{d_d}{d_b + d_a}\right)^r$. The resampling algorithm described in Algorithm 1 is applied to the data points after the weights have been normalized to sum to unity $w_n^r = w^r / \sum_{r=1}^{R} w^r$.

Algorithm 1 *Given R particles/data points X_n^r and their weights w_n^r compute new set of particles/data points X_n^s and equal weights w_n^s*

1. Initialize the cumulative sum of weights (CW) $c_1 = w_n^1$
for $r = 2 : R$ **do**
construct CW: $c_r = c_{r-1} + w_n^r$
end for
2. Start at the base of CW: $r = 1$
3. Draw a starting point $\mu_1 = \{0, R^{-1}\}$
for $s = 1 : R$ **do**
Move along CW $\mu_s = \mu_1 + R^{-1}(s-1)$
while $(\mu_s > c_r)$ **do**
$r = r + 1$;
end while
4. Reassign particles : $X_n^s = X_n^r$
5. Reassign weight : $w_n^s = N^{-1}$
6. Reassign parents : $r^s = r$
end for

end

In our application of the resampling algorithm, the number of input data points and output data points can be changed. In fact better ellipse fitting results are obtained when the number of output data points exceeds the number of input data points. We adopted the strategy of keeping all data points and replicating those points having higher weights. For replication the data points were interpolated between current parent point and the previous neighbour of the point. The number of replication is a function of the weight of the parent point. To compute the weights of data points a knowledge of the major and minor axis is required. In the experiment with real data set for the example shown in Sect. 14.8 and Fig. 14.11, an estimate of the major and minor axes is obtained by taking means of the ellipse parameters obtained by the methods used for ellipse fitting. The estimation of the eccentric weight of data points is not sensitive to small errors in the estimate of minor and major axes.

14.7 Camera Self Calibration

The value of the image of the rotation axis $\mathbf{l}_s$ is initialized by the line which best fits the poles of the ellipses fitted to the trajectories of the root tips. In a real perspective image of circular motion it is not possible to have four real intersection points for

the conics fitted to the trajectories of the feature points. When a pair of conics have two real intersection points then the other pair of complex conjugate intersection points are the image of the circular points $\mathbf{i}, \mathbf{j}$. Let say $\mathbf{i} = [a + ib, c - id, 1]^T$ then $\mathbf{j} = [a - ib, c + id, 1]^T$. The line passing through the circular points is the image of the horizon $\mathbf{l}_h$. Thus

$$\mathbf{l}_h = \mathbf{i} \times \mathbf{j} \tag{14.8}$$

Complex conjugate intersection of each pair of conics gives an estimate of a, b, c, d which specifies the circular points. If there are Q number of conics in a circular motion sequence then there will be $\binom{Q}{2}$ estimates of a, b, c, d. In our approach we histogram the computed values of a, b, c, d and take the modes of the histogram to be the estimates of a, b, c, d and hence of the circular points $\mathbf{i}, \mathbf{j}$. From the estimates of the image invariants of the single axis motion and the feature point tracking we have to derive the parameters of internal calibration of the camera the $\mathbf{K}$ matrix and the poses of the camera which will be used to reconstruct the visual hull of the roots from its silhouettes.

14.7.1 Internal Calibration

To estimate the internal parameters of the camera we assume the camera to have zero skew and its pixels to be square. For a camera with zero skew and square pixels the image of absolute conics (IAC) $\boldsymbol{\omega}$ is of the following form [17]

$$\boldsymbol{\omega} = \begin{bmatrix} \omega_1 & 0 & \omega_2 \\ 0 & \omega_1 & \omega_3 \\ \omega_2 & \omega_3 & \omega_4 \end{bmatrix} \tag{14.9}$$

and the matrix for internal parameters of the camera is

$$\mathbf{K} = \begin{bmatrix} f & 0 & x_0 \\ 0 & f & y_0 \\ 0 & 0 & 1 \end{bmatrix}. \tag{14.10}$$

The pair of circular points constraints $\boldsymbol{\omega}$ by the following equality $\mathbf{i}\boldsymbol{\omega}\mathbf{i}^T = \mathbf{j}\boldsymbol{\omega}\mathbf{j}^T = 0$. This constraint is not sufficient to compute all the elements of $\boldsymbol{\omega}$ as $\mathbf{i}$ and $\mathbf{j}$ are complex conjugates of each other. The added constraint from the vanishing points $\mathbf{v}_z^T \boldsymbol{\omega} \mathbf{v}_y = \mathbf{v}_y^T \boldsymbol{\omega} \mathbf{v}_x = \mathbf{v}_x^T \boldsymbol{\omega} \mathbf{v}_z = 0$ and the constraint from the pole polar relationship $\mathbf{l}_s = \boldsymbol{\omega}\mathbf{v}_x$ and $\mathbf{l}_h = \boldsymbol{\omega}\mathbf{v}_y$ are sufficient to compute $\boldsymbol{\omega}$ in a least square sense using SVD. The relationship of $\boldsymbol{\omega}$ to the internal parameters of a camera $\mathbf{K}$ is $\boldsymbol{\omega} = (\mathbf{K}\mathbf{K}^T)^{-1}$ [17, 51]. By applying Cholesky decomposition to $\boldsymbol{\omega}^{-1}$ [17]

one can obtain an upper triangular matrix for $\mathbf{K}$ of the from $\begin{bmatrix} k_{11} & k_{12} & k_{13} \\ 0 & k_{22} & k_{23} \\ 0 & 0 & k_{33} \end{bmatrix}$. In carrying out the Cholesky decomposition we experienced that usually $k_{12} \neq 0$ and $k_{11} \neq k_{22}$, which is the desirable form for $\mathbf{K}$ in Eq. (14.10), for a zero skew square pixel camera. When $\mathbf{K}$ is of the form in Eq. (14.10) then $(\mathbf{K}\mathbf{K}^T)^{-1}$ is of the from

$$(\mathbf{K}\mathbf{K}^T)^{-1} = \frac{1}{f^4}\begin{pmatrix} f^2 & 0 & -f^2x_0 \\ 0 & -f^2 & -f^2y_0 \\ -f^2x_0 & -f^2y_0 & f^4 + f^2x_0^2 + f^2y_0^2 \end{pmatrix}. \tag{14.11}$$

In our implementation we solved for f, x_0, y_0 by equating the elements of the above matrix to that of element of $\boldsymbol{\omega}$ as in Eq. (14.9).

14.7.2 Camera Pose Estimation

A projective camera matrix $\mathbf{P}$ transform a point in 3D to a point in the image plane and is of the form

$$\mathbf{P} = \mathbf{K}\mathbf{R}[\mathbf{I}|\mathbf{t}] \tag{14.12}$$

For the geometry of the reference camera C_0, the projection matrix is of the form

$$\mathbf{P}_0 = \mathbf{K}\mathbf{R}\begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & t_z \end{bmatrix} \tag{14.13}$$

where t_z is the location of the camera centre on the Z-axis and $\mathbf{R}$ is the rotation matrix of the form $(\mathbf{r}_1, \mathbf{r}_2, \mathbf{r}_3)$, which describes the initial orientation of the camera [48]. Its columns can be recovered as

$$\mathbf{r}_1 = \frac{\mathbf{K}^{-1}\mathbf{v}_x}{|\mathbf{K}^{-1}\mathbf{v}_x|}, \mathbf{r}_3 = \frac{\mathbf{K}^{-1}\mathbf{v}_z}{|\mathbf{K}^{-1}\mathbf{v}_z|}, \text{ and } \mathbf{r}_2 = \mathbf{r}_1 \times \mathbf{r}_3. \tag{14.14}$$

The projection matrix of camera rotated from the reference camera by θ_i is given by

$$\mathbf{P}_i = \mathbf{K}\mathbf{R}\begin{bmatrix} \cos\theta_i & 0 & \sin\theta_i & 0 \\ 0 & 1 & 0 & 0 \\ -\sin\theta_i & 0 & \cos\theta_i & t_z \end{bmatrix} \tag{14.15}$$

To estimate θ_i from the image features we applied Laguerre's formula as used in [23] and computed from the trajectories of the feature points.

14.8 Results

In our classification experiments the principle order n_{max} of the Zernike moments is varied from 2 through to 12 in steps of 2 for a fixed number of Gaussians K in the GMM density model. The order 0 and 1 are omitted due to the nature of preprocessing applied in our experiments. The lower order moments capture the overall (generic) shape information, while the higher order moments capture the detailed (specific) shape information. The value of η_{th} is varied from 0.0 to 900,000 and each point on the ROC curve is the True detection rate and False alarm rate computed for a particular value of η_{th}. Figure 14.7 shows the plot of the ROC curves for $K = 4$ and different values of the Zernike moments.

Figure 14.8 shows the ROC curves for $n_{max} = 8$ and $K = \{2, 4, 6, 8, 10\}$. In this plot we observe that the classification results improves with initial increase in the number of Gaussians K in the GMM for the class conditional densities. The best ROC curve is for $K = 6$ with the quality of the curve deteriorating with further increase in the value of K. To find out which combination of the principal order of the Zernike moments, n_{max}, and the number of Gaussians in the GMM, K, gives the best classification results we plotted the area under (AU) the ROC curves as a function of K for different values of n_{max}. This plot is shown in Fig. 14.9.

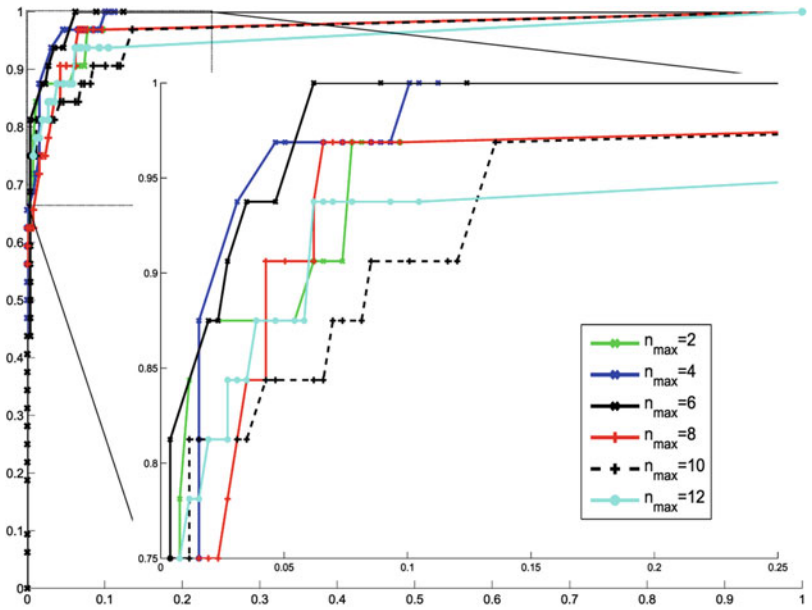


Fig. 14.7 ROC curves for the classification results using the proposed classification method. The number of Gaussians K in the GMM is 4 for principle Zernike moments $n_{max} = \{2, 4, 6, 8, 10, 12\}$

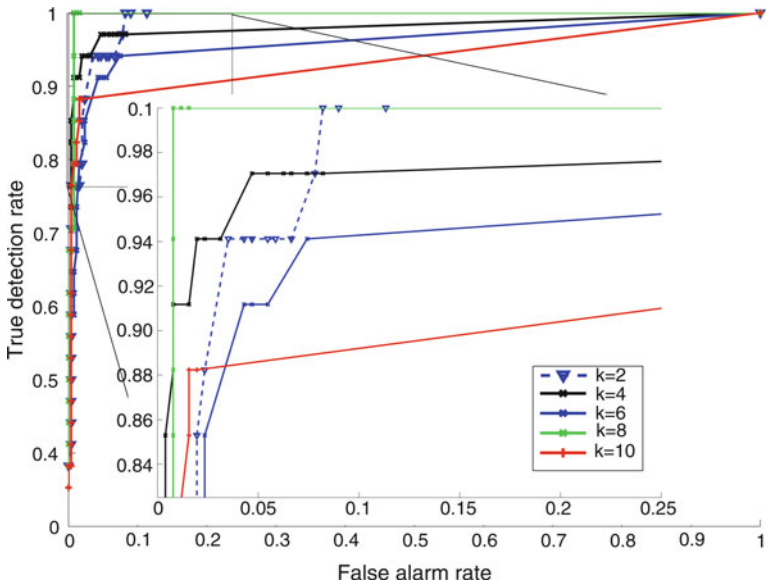


Fig. 14.8 ROC curves for the classification results using the proposed classification method. The principal Zernike moments $n_{max} = 8$ and the number of Gaussians K in the GMM are $\{2, 4, 6, 8, 10\}$

The highest peaks in this graph occur for the combinations of $\{n_{max}, K\} = \{6, 8\}$ and $\{n_{max}, K\} = \{8, 6\}$, respectively, giving the best performance combinations. The areas under the ROC curves for these combinations are 0.9972. As K is further increased, the performance of classification deteriorates and the convergence of the EM algorithm becomes unstable for high values of K. The high values of K we can experiment with also depends upon the size of the training dataset. At present our dataset consists of about 1,500 non-root tip patches and 500 root tip patches.

To quantitatively compare our results with support vector machine (SVM) classification results. Using the same data set an SVM was trained for classification where the parameters of the SVM has been optimized using cross validation. We use the same feature, Zernike moments in SVM as the choice of feature may affect the classification results and hence the comparison will not be a valid one if different feature is used for classification using SVM. Figure 14.10 show the ROC curves for the SVM classification results for different values of n_{max} Here the true detection rate (TDR) of 100 % is was achieved at the false detection rate (FDR) of 37 % for $n_{max} = 2$. The classification rate for lower n_{max} are better than for the higher values of n_{max}.

To investigate the effectiveness of the ellipse fitting method we show some qualitative results on real data (for quantitative evaluation please refer to [35]).

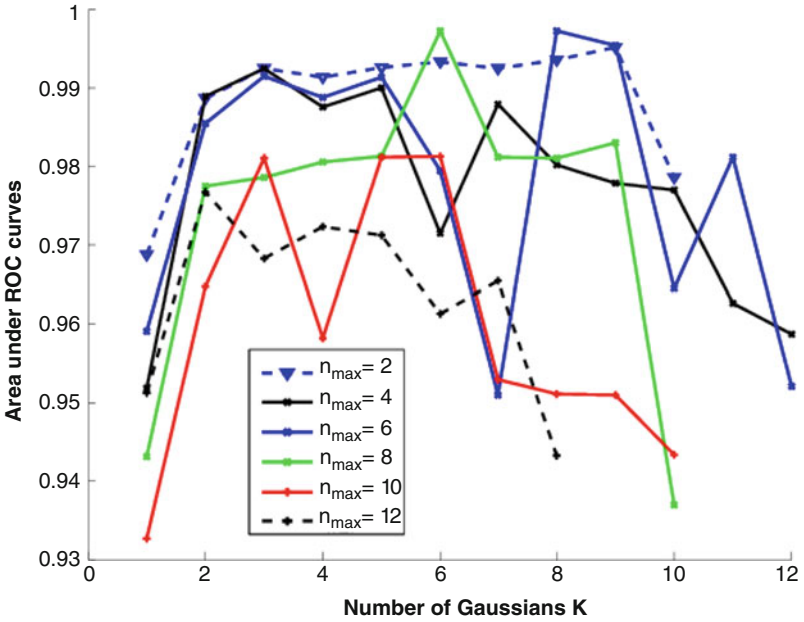


Fig. 14.9 Plot of the area under the curve for the classification results using the proposed method. Best performance is achieved for the combination principal Zernike moment order $n_{max} = \{6, 8\}$ and number of Gaussian in the GMM $K = \{8, 6\}$. The area under the ROC for this combination are 0.9972

Resampling was applied as a precursor to the five different ellipse fitting methods: CGIP-1979 [5], PAMI-1999 [13], WSCG-1998 [15], PAMI-1991 [44], and ECCV-2012 [43]. The data set used for qualitative testing the proposed algorithm comes from trajectories of root tips. The trajectory of a single root tip traces out an ellipse in the image plane. Ellipse fitting is then applied to the set of detected root tips in the 72 different positions. Figure 14.11 shows the results of this fitting. The small regions (a), (b), (c) of Fig. 14.11 are magnified for better visualization. The left hand subplots are the results of fitting before resampling, while the right hand subplots show the results after resampling. There is significant improvement in the results of ellipse fitting after resampling of data points. An arrow marker is inserted in the images to highlight the improvement in the results, especially for the PAMI-1999 method.

In the results we show the 3D reconstruction and modelling of a corn root at its different stages of development and also when different number of images are used for reconstruction of the root. Figure 14.12 shows the reconstruction of the RSA at: day 6 (a), day 10 (b), day 14 (c), and day 18 (d). These reconstructions are

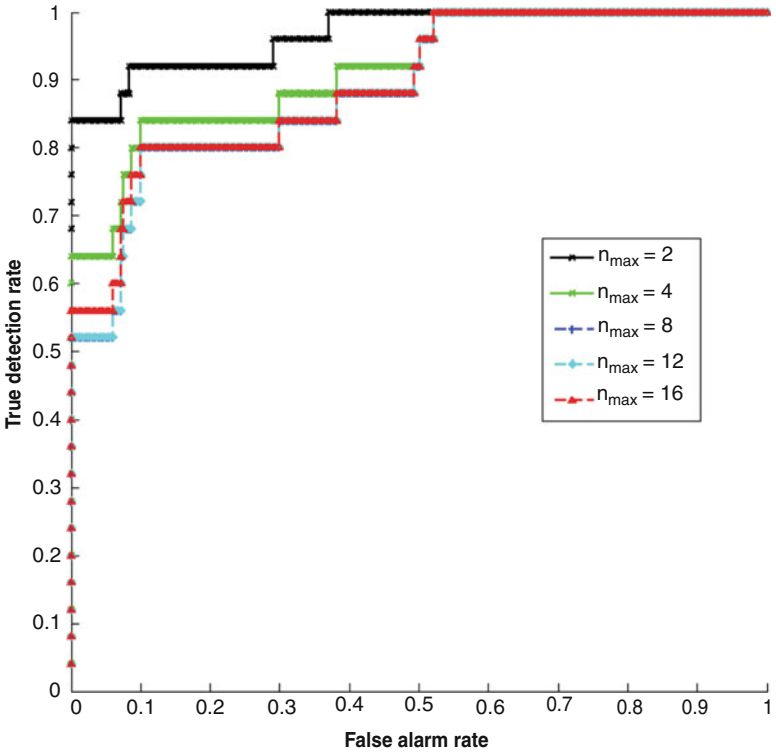


Fig. 14.10 ROC curves for the classification results using the SVM cross validation method for different values of the principle Zernike moments $n_{max} = \{2, 4, 8, 12, 16\}$. Better ROC curves are obtained for lower values of n_{max}

sufficiently detailed and truly portrays the complex 3D structure of RSA. Models in these roots have been developed using all 72 images from the turntable image sequence. Figure 14.13 shows the reconstruction of the root with different number of images used for 3D reconstruction: 12 images Fig. 14.13a, 24 images Fig. 14.13b, 36 images Fig. 14.13c, 48 images Fig. 14.13d, 60 images Fig. 14.13e and 72 images Fig. 14.13f. The results are when the plant is at day 14th of its growth. There are subtle differences in these 3D models of the root. As the number of images in the 3D reconstruction are increased the fine structures of the RSA are captured and represented in the 3D model. This can be noted by observing the longest seminal root in these models. This seminal root is longest and more vividly captured in Fig. 14.13f than the other images. The stoutness of the models from Fig. 14.13a–f seems to decrease.

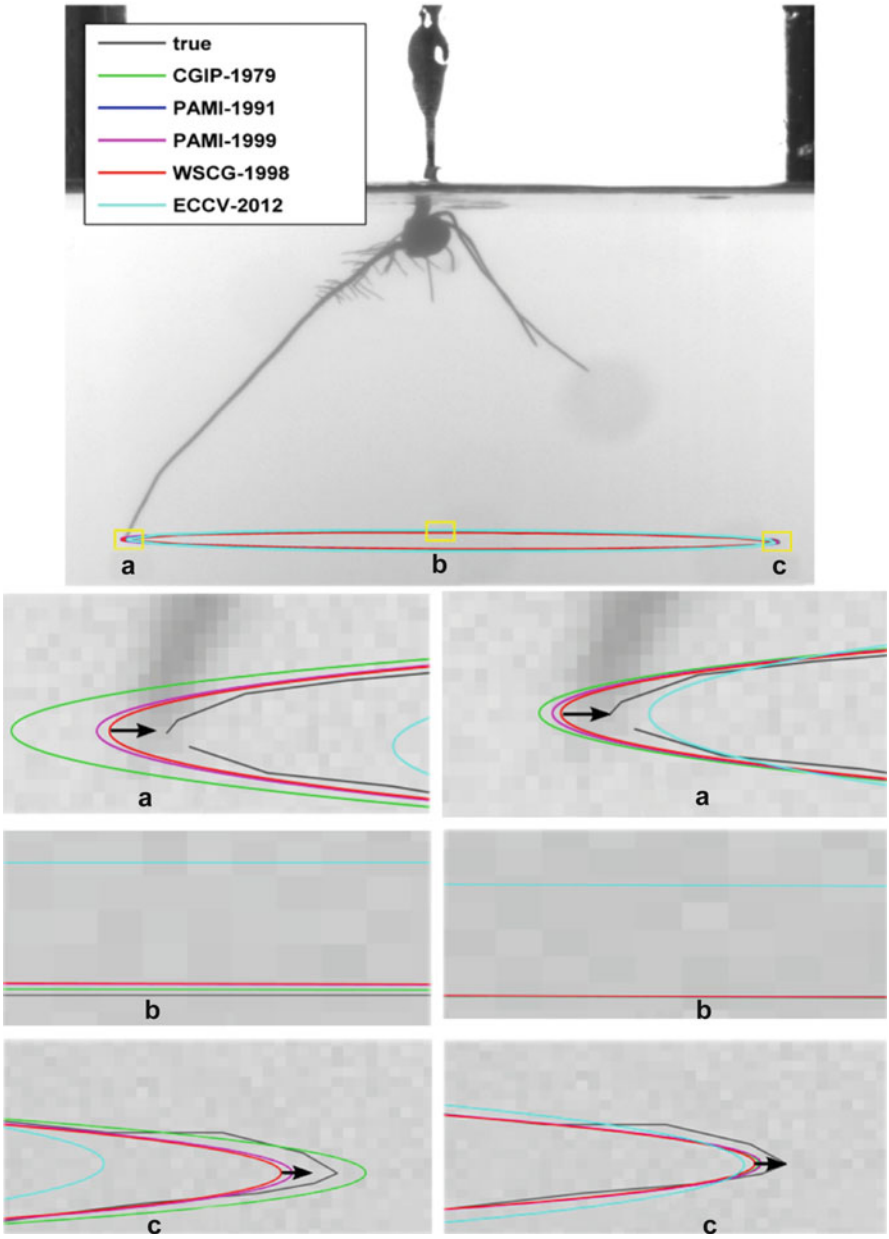


Fig. 14.11 Results of ellipse fitting of a real data set after resampling. The main figure shows first the origin (root tip tracking) of a real data set, plus overall results of ellipse fitting. Subplots show magnified views of small regions (*a*), (*b*), and (*c*). *Left-hand* subplots are results prior to resampling. *Right-hand* subplots are results subsequent to resampling. An equal length *arrow* is included in both sets to highlight the obvious improvements

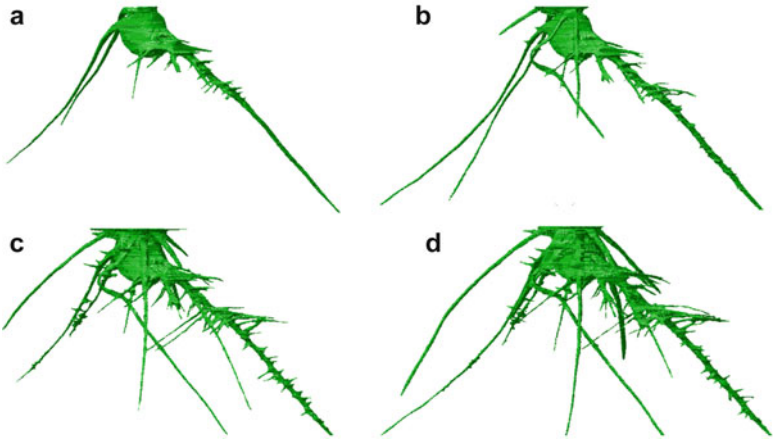


Fig. 14.12 (**a**, **b**, **c**, **d**) are the results of 3D reconstruction for day 6, 10, 14, 18 of the growth of corn roots. The results by the proposed method are detailed enough for phenotyping analysis (Images courtesy of [36])

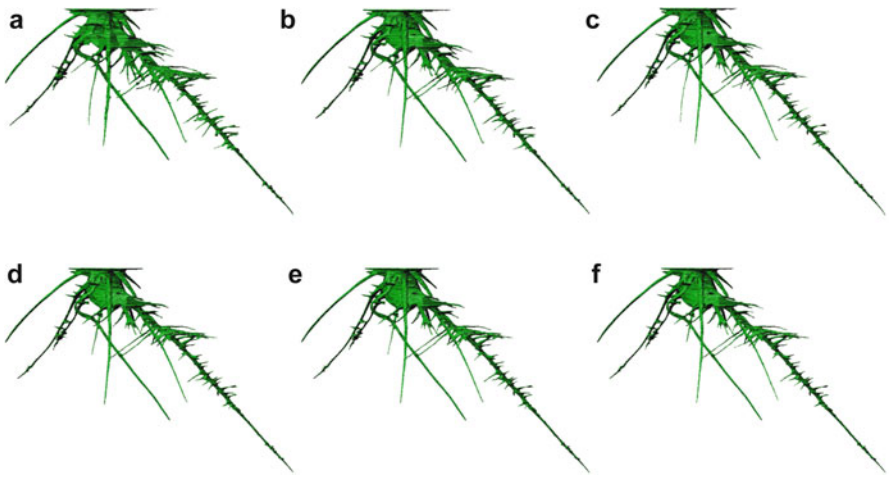


Fig. 14.13 (**a**, **b**, **c**, **d**, **e**, **f**) are the results of 3D reconstruction with 12, 24, 36, 48, 60, 72 number of images respectively, from the turntable sequence. The images are from the day 14 growth stage of the corn plant (Images courtesy of [36])

Conclusions

In this chapter we have presented a complete system for 3D reconstruction and analysis of roots grown in transparent gellan gum medium. Although not a natural environment for growing plants, never the less it represents a useful strategy for phenotyping roots and studying the 3D structure using relatively in-expensive techniques. Visual sensors (digital) cameras are much cheaper to procure, operate, and maintain than 3D CT X-ray scanners, MRI, or even Lidar. We have demonstrated the successful working of the proposed method on corn roots. Detailed reconstruction of the roots where obtained with 72 images. The root growth rate is observed to be almost linear with time, which is well in line with current theories for plants in biology that is a plant in its early stage of growth has almost a linear growth rate of biomass. Increasing the number of images for 3D reconstruction improves the accuracy of 3D model although it decreases the observed volume. This empirical finding is consistent with the theory of space carving which states that the volume of reconstruction decreases monotonically with increasing number of images from different view points. In our future experiments we will also consider the effect of changing viewing angles. We will also report on the issues due to refraction of light through gellan gum and its impact on 3D reconstruction.

References

1. S.J. Ahn, W. Rauh, H.-J. Warnecke, Least-squares orthogonal distances fitting of circle, sphere, ellipse, hyperbola, and parabola. Pattern Recognit. **34**(12), 2283–2303 (2001)
2. S.J. Ahn, W. Rauh, H.S. Cho, H.-J. Warnecke, Orthogonal distance fitting of implicit curves and surfaces. IEEE Trans. Pattern Anal. Mach. Intell. **24**(5), 620–638 (2002)
3. S. Arulampalam, S. Maskell, N. Gordon, T. Clapp, A tutorial on particle filters for on-line non-linear/non-Gaussian Bayesian tracking. IEEE Trans. Signal Process. **50**(2), 174–188 (2002)
4. X. Bai, C. Sun, F. Zhou, Splitting touching cells based on concave points and ellipse fitting. Pattern Recognit. **42**(11), 2434–2446 (2009)
5. F.L. Bookstein, Fitting conic sections to scattered data. Comput. Graph. Image Process. **9**, 56–71 (1979)
6. X. Cao, J. Xiao, H. Foroosh, M. Shah, Self-calibration from turn-table sequences in presence of zoom and focus. Comput. Vis. Image Underst. **102**(3), 227–237 (2006)
7. C.-C. Chang, C.-J. Lin, LIBSVM: a library for support vector machines. ACM Trans. Intell. Syst. Technol. **2**, 27:1–27:27 (2011)
8. W. Chojnacki, M.J. Brooks, A. van den Hengel, D. Gawley, On the fitting of surfaces to data with covariances. IEEE Trans. Pattern Anal. Mach. Intell. **22**, 1294–1303 (2000)
9. R.T. Clark, R.B. MacCurdy, J.K. Jung, J.E. Shaff, S.R. McCouch, D.J. Aneshansley, L.V. Kochian, Three-dimensional root phenotyping with a novel imaging and software platform. Plant Physiol. **156**(2), 455–465 (2011)
10. I. De Smet, P.J. White, A.G. Bengough, L. Dupuy, B. Parizot, I. Casimiro, R. Heidstra, M. Laskowski, M. Lepetit, F. Hochholdinger, X. Draye, H. Zhang, M.R. Broadley, B. Péret, J.P. Hammond, H. Fukaki, S. Mooney, J.P. Lynch, Ph. Nacry, U. Schurr, L. Laplaze, P. Benfey, T. Beeckman, M. Bennett, Analyzing lateral root development: How to move forward. Plant Cell **24**(1), 15–20 (2012)

11. F. Duan, L. Wang, P. Guo, Ransac based ellipse detection with application to catadioptric camera calibration, in *ICONIP 2010*, Sydney, pp. 525–532
12. A.W. Fitzgibbon, G. Cross, A. Zisserman, Automatic 3D model construction for turn-table sequences, in *Proceedings of the European Workshop on 3D Structure from Multiple Images of Large-Scale Environments, SMILE'98*, Freiburg (Springer, London, 1998), pp. 155–170
13. A. Fitzgibbon, M. Pilu, R.B. Fisher, Direct least square fitting of ellipses. IEEE Trans. Pattern Anal. Mach. Intell. **21**(5), 476–480 (1999)
14. K. Forbes, F. Nicolls, G. De Jager, A. Voigt, Shape-from-silhouette with two mirrors and an uncalibrated camera, in *Proceedings of the 9th European Conference on Computer Vision (ECCV)*, Graz, 2006, pp. 165–178
15. R. Halir, J. Flusser, Numerically stable direct least squares fitting of ellipses, in *Proceedings of the 6th International Conference in Central Europe on Computer Graphics and Visualization WSCG'98*, Plzen, 1998, pp. 125–132
16. G. Hammer, Z. Dong, G. Mclean, A. Doherty, C. Messina, J. Schussler, C. Zinselmeier, S. Paszkiewicz, M. Cooper, Can changes in canopy and/or root system architecture explain historical maize yield trends in the U.S. corn belt? Crop Sci. **49**, 299–312 (2009)
17. R.I. Hartley, A. Zisserman, *Multiple View Geometry in Computer Vision*, 2nd edn. (Cambridge University Press, 2004). ISBN: 0521540518
18. X.C. He, N.H.C. Yung, Corner detector based on global and local curvature properties. Opt. Eng. **47**(5), 057008 (2008)
19. C. Hernandez, F. Schmitt, R. Cipolla, Silhouette coherence for camera calibration under circular motion. IEEE Trans. Pattern Anal. Mach. Intell. **29**(2), 343–349 (2007)
20. A. Hodge, G. Berta, C. Doussan, F. Merchan, M. Crespi, Plant root growth, architecture and function. Plant Soil **321**(1–2), 153–187 (2009)
21. A.K. Jain, S. Prabhakar, L. Hong, S. Pankanti, Filterbank-based fingerprint matching. IEEE Trans. Image Process. **9**, 846–859 (2000)
22. G. Jiang, L. Quan, Circular motion geometry using minimal data. IEEE Trans. Pattern Anal. Mach. Intell. **26**, 721–731 (2004)
23. G. Jiang, H.-T. Tsui, L. Quan, A. Zisserman, Single axis geometry by fitting conics, in *Computer Vision – ECCV 2002*, Copenhagen, ed. by A. Heyden, G. Sparr, M. Nielsen, P. Johansen. Volume 2350 of Lecture Notes in Computer Science (Springer, Berlin/Heidelberg, 2002), pp. 537–550
24. G. Jiang, L. Quan, H.-T. Tsui, Circular motion geometry by minimal 2 points in 4 images, in *Ninth IEEE International Conference on Computer Vision*, Nice, vol. 1, 2003, pp. 221–227
25. G. Jiang, H.-T. Tsui, L. Quan, A. Zisserman, Geometry of single axis motions using conic fitting. IEEE Trans. Pattern Anal. Mach. Intell. **25**(10), 1343–1348 (2003)
26. K. Kanatani, Ellipse fitting with hyperaccuracy. IEICE Trans. Inf. Syst. **89**, 2653–2660 (2006)
27. K. Kanatani, P. Rangarajan, Hyperaccurate ellipse fitting without iterations, in *VISAPP (2)*, Angers, 2010, pp. 5–12
28. A. Khotanzad, Y.H. Hong, Invariant image recognition by Zernike moments. IEEE Trans. Pattern Anal. Mach. Intell. **12**, 489–497 (1990)
29. P. Kumar, K. Sengupta, A. Lee, A comparative study of different color spaces for foreground and shadow detection for traffic monitoring system, in *IEEE 5th International Conference on Intelligent Transportation Systems*, Singapore (IEEE, 2002), pp. 100–105
30. P. Kumar, S. Ranganath, W.M. Huang, Queue based fast background modelling and fast hysteresis thresholding for better foreground segmentation, in *The Fourth International Conference on Information, Communications and Signal Processing*, Singapore, vol. 2, 2003, pp. 743–747
31. P. Kumar, S. Ranganath, K. Sengupta, W. Huang, Co-operative multi-target tracking and classification, in *Computer Vision – ECCV 2004*, Prague, ed. by T. Pajdla, J. Matas. Volume 3021 of Lecture Notes in Computer Science (Springer, Berlin/Heidelberg, 2004), pp. 376–389
32. P. Kumar, S. Ranganath, W.M. Huang, K. Sengupta, Framework for real time behavior interpretation from traffic video. IEEE Trans Intell Transp. Syst. **6**(1), 43–53 (2005)

33. P. Kumar, S. Ranganath, K. Sengupta, W. Huang, Cooperative multitarget tracking with efficient split and merge handling. IEEE Trans. Circuits Syst Video Technol. **16**(12), 1477–1490 (2006)
34. P. Kumar, J. Cai, S. Miklavcic, Automated detection of root crowns using Gaussian mixture model and bayes classification, in *International Conference on Digital Image Computing Techniques and Applications (DICTA)*, Fremantle, 2012, pp. 1–7
35. P. Kumar, J. Cai, S. Miklavcic, Improved ellipse fitting by considering the eccentricity of data point sets, in *Proceedings of IEEE International Conference on Image Processing (ICIP)*, Melbourne, 2013, pp. 815–819
36. P. Kumar, J. Cai, S. Miklavcic, 3D reconstruction, modelling and analysis of in situ root system architecture, in *20th International Congress on Modelling and Simulation*, Adelaide, 1–6 Dec 2013, pp. 517–523
37. G. Lobet, L. Pagès, X. Draye, A novel image-analysis toolbox enabling quantitative analysis of root system architecture. Plant Physiol. **157**(1), 29–39 (2011)
38. J. Lynch, Root architecture and plant productivity. Plant Physiol. **109**, 7–13 (1995)
39. P.R.S. Mendonça, K.-Y.K. Wong, R. Cipolla, Epipolar geometry from profiles under circular motion. IEEE Trans. Pattern Anal. Mach. Intell. **23**, 604–616 (2001)
40. M.A. Moreno-Risueno, J.M. Van Norman, A. Moreno, J. Zhang, S.E. Ahnert, P.N. Benfey, Oscillating gene expression determines competence for periodic arabidopsis root branching. Science **329**(5997), 1306–1311 (2010)
41. M. Sarzi, H. Rix, J.C. Sheilds, G. Rudnick, D.H. McIntosh L.C. Ho, A.V. Filippenko, W.L.W. Sargent, Supermassive black holes in bulges. Astrophys. J. **550**(1), 65–74 (2001)
42. S.N. Sinha, M. Pollefeys, L. McMillan, Camera network calibration from dynamic silhouettes, in *Proceedings of the 2004 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, Washington, DC, vol. 1, 27 June–2 July 2004, pp. I-195–I-202
43. Z.L. Szpak, W. Chojnacki, A. van den Hengel, Guaranteed ellipse fitting with the sampson distance, in *ECCV (5)*, Florence, 2012, pp. 87–100
44. G. Taubin, Estimation of planar curves, surfaces, and nonplanar space curves defined by implicit equations with applications to edge and range image segmentation. IEEE Trans. Pattern Anal. Mach. Intell. **13**(11), 1115–1138 (1991)
45. M.R. Teague, Image analysis via the general theroy of moments. J. Opt. Soc. Am. **70**(8), 920–930 (1980)
46. J. Traas, T. Vernoux, Oscillating roots. Science **329**(5997), 1290–1291 (2010)
47. A. Vedaldi, B. Fulkerson, VLFeat: an open and portable library of computer vision algorithms (2008). http://www.vlfeat.org/
48. K.-Y.K. Wong, G. Zhang, C. Liang, H. Zhang, 1D camera geometry and its application to the self-calibration of circular motion sequences. IEEE Trans. Pattern Anal. Mach. Intell. **30**(12), 2243–2248 (2008)
49. X. Ying, Z. Hu, Catadioptric camera calibration using geometric invariants. IEEE Trans. Pattern Anal. Mach. Intell. **26**, 1260–1271 (2004)
50. J. Yu, S.R. Kulkarni, H.V. Poor, Robust ellipse and spheroid fitting. Pattern Recognit. Lett. **33**(5), 492–499 (2012)
51. H. Zhang, K.-Y.K. Wong, Self-calibration of turntable sequences from silhouettes. IEEE Trans. Pattern Anal. Mach. Intell. **31**(1), 5–14 (2009)
52. G. Zhang, D.S. Jayas, N.D. White, Seperation of touching grain kernels in an image by ellipse fitting algorithm. Biosyst. Eng. **92**(2), 135–142 (2005)
53. H. Zhang, L. Shao, K.-Y.K. Wong, Self-calibration and motion recovery from silhouettes with two mirrors, in *Proceedings of the Asian Conference on Computer Vision*, Daejeon, 2012, pp. 1–12
54. Y. Zheng, S. Gu, H. Edelsbrunner, C. Tomasi, P. Benfey, Detailed reconstruction of 3D plant root shape, in *ICCV*, Barcelona, 2011, pp. 2026–2033

Index

C. Sun et al. (eds.), *Signal and Image Analysis for Biomedical and Life Sciences*,
Advances in Experimental Medicine and Biology 823,
DOI 10.1007/978-3-319-10984-8