University of Technology, Sydney

*Doctoral Thesis*

# Ensemble Rapid Centroid Estimation: A Semi-Stochastic Consensus Particle Swarm Approach for Large Scale Cluster Optimization

*Supervisors:*
A/Prof. Steven W. Su
Dr. Bruce D. Moulton
Prof. Hung T. Nguyen

*Author:*
Mitchell Yuwono

*A thesis submitted in fulfilment of the requirements for the degree of*

*Doctor of Philosophy*

*in the*

Centre for Health Technologies
School of Electrical, Mechanical and Mechatronic Systems
Faculty of Engineering and Information Technology

February 2015

# Declaration of Authorship

I, Mitchell Yuwono, declare that this thesis titled, 'Ensemble Rapid Centroid Estimation: A Semi-Stochastic Consensus Particle Swarm Approach for Large Scale Cluster Optimization' and the work presented in it are my own. I confirm that:

- This work was done wholly or mainly while in candidature for a research degree at this University.
- Where any part of this thesis has previously been submitted for a degree or any other qualification at this University or any other institution, this has been clearly stated.
- Where I have consulted the published work of others, this is always clearly attributed.
- Where I have quoted from the work of others, the source is always given. With the exception of such quotations, this thesis is entirely my own work.
- I have acknowledged all main sources of help.
- Where the thesis is based on work done by myself jointly with others, I have made clear exactly what was done by others and what I have contributed myself.

Signed: ________________________________

Date: ________________________________

*“Start by doing what’s necessary; then do what’s possible; and suddenly you are doing the impossible.”*

St. Francis of Assisi

UNIVERSITY OF TECHNOLOGY, SYDNEY

# *Abstract*

Faculty of Engineering and Information Technology
School of Electrical, Mechanical and Mechatronic Systems

Doctor of Philosophy

**Ensemble Rapid Centroid Estimation: A Semi-Stochastic Consensus Particle Swarm Approach for Large Scale Cluster Optimization**

by Mitchell Yuwono

This thesis details rigorous theoretical and empirical analyses on the related works in the clustering literature based on the Particle Swarm Optimization (PSO) principles. In particular, we detail the discovery of disadvantages in Van Der Merwe - Engelbrecht's PSO clustering, Cohen - de Castro Particle Swarm Clustering (PSC), Szabo's modified PSC (mPSC) and Szabo's Fuzzy PSC (FPSC).

We validate, both theoretically and empirically, that Van Der Merwe - Engelbrecht's PSO clustering algorithm is not significantly better than the conventional $k$-means. We propose that under random initialization, the performance of their proposed algorithm diminishes exponentially as the number of classes or dimensions increase.

We unravel that the PSC, mPSC, and FPSC algorithms suffer from significant complexity issues which do not translate into performance. Their cognitive and social parameters have negligible effect to convergence and the algorithms generalize to the $k$-means, retaining all of its characteristics including the most severe: the curse of initial position. Furthermore we observe that the three algorithms, although proposed under varying names and time frames, behave similarly to the original PSC.

This thesis analyzes, both theoretically and empirically, the strengths and limitations of our proposed semi-stochastic particle swarm clustering algorithm, Rapid Centroid Estimation (RCE), self-evolutionary Ensemble RCE (ERCE), and Consensus Engrams, which are developed mainly to address the fundamental issues in PSO Clustering and the PSC families. The algorithms extend the scalability, applicability, and reliability of earlier approaches to handle large-scale non-convex cluster optimization in quasilinear complexity in both time and space. This thesis establishes the fundamentals, much surpassing those outlined in our published manuscripts.

# *Acknowledgements*

First and foremost I would like to express my gratitude to my supervisors, A/Prof. Steven Su, Dr. Bruce D. Moulton, and Prof Hung T. Nguyen. Their vast knowledge, intelligence, expertise, understanding, and patience greatly expanded my knowledge and skill in many areas in the course of my PhD research. I really appreciate the considerable amount of efforts to help me with difficult research projects, critical reviews of my writings, and ultimately my scholarship applications. They are exceptional, inspirational, and excellent supervisors. Under their supervision, we publish high quality papers in prestigious journals.

A very special thanks goes to Prof. Hung T. Nguyen, who has been a visionary since the beginning of my Masters studies. Prof. Nguyen has inspired me to dive deeper into the world of biomedical technologies. It was under his supervision that I developed a focus and became interested in biomedical research and dared to explore the world of applied machine learning.

I would like to express my most sincere gratitude to my principal supervisor, A/Prof. Steven W. Su, who has been with me since the start of my PhD and has been through my ups and downs in my PhD life. He has seen through my struggles, supervised through numerous difficult research projects, and ultimately mend me to be an efficient, independent and resilient researcher. A/Prof. Steven is a kindhearted and understanding supervisor who is always supportive at all times.

I would also like to thank Dr. Bruce D. Moulton for being very critical on all aspects of my research life, especially with regards to the correctness of experiments, reports, and journals. His warm personality, deep knowledge and insights, and also attention to details are sincerely worth admiring. The works that I have done under his supervision have always been brought to perfection.

I would also like to thank to Dr. Jaime Valls Miro that has supervised me through the course of my Masters project in Advanced Robotics and supervised me during the summer research internship in ARC Centre for Autonomous Systems in UTS. The depth of his knowledge and his warm character have been a great inspiration throughout my life as a graduate student. Until today, Dr. Jaime has always been a role model in my life as a researcher.

I would also like to thank Dr. Ying Guo, Dr. Glenn Platt, Dr. Josh Wall, and Mr. Sam West from The Commonwealth Scientific and Industrial Research Organization (CSIRO) for the opportunity to apply our research in a practical industrial setting.

The members of the Centre for Health Technologies (CHT) have contributed to my personal and professional time at UTS. The group has been a source of friendships as well as good advice and collaboration. I am especially grateful to Father Ardi Handojoseno, who has been a source of inspiration, fun, moral and spiritual support during the course of my graduate studies, the up and downs in my Masters and PhD. I also would like to thank Phyo Phyo San; Henry Chandra; Rifai Chai; Andrew Zhang; Bao C. Q. Truong; Prof. Steve Ling; Ganesh Naik; Gibson Hu; Shuo Wang; Lei Zhang; and all of my colleagues whose name has not been mentioned here.

I would like to thank mom and dad for being the greatest parents in the universe. There is no words capable of conveying my feelings of gratitude towards the both of them. This thesis wouldn't exist if it is not because of their love. This thesis is my gift for them.

I would like to sincerely thank the love of my life. She has been there since the beginning of the writing until the completion of this thesis. Her contribution is enormous, both in this thesis and in my life as a whole.

Finally I would like to express my deepest gratitude to the sponsors for my scholarships including: University of Technology Sydney for their financial supports through the International Research Scholarship (IRS) and UTS President Scholarship (UTSP); The Commonwealth Scientific and Industrial Research Organization (CSIRO) for the scholarship and additional projects. I recognize that this research would not have been possible without their supports.

Mitchell Yuwono
*University of Technology, Sydney*
September 2014

# Contents

**Declaration of Authorship** **ii**

**Abstract** **iv**

**Acknowledgements** **v**

**Contents** **vii**

**List of Figures** **xi**

**List of Tables** **xv**

**Symbols** **xvii**

**List of Publications** **1**

**1 Introduction** **3**
- 1.1 Overview . . . 4
- 1.2 Clustering in Biomedical Informatics . . . 5
- 1.3 Applications . . . 5
- 1.4 Challenges . . . 12
- 1.5 Contribution of This Thesis . . . 13

**2 Literature Review** **25**
- 2.1 Challenges in Cluster Analysis . . . 25
- 2.2 Quantifying Dissimilarity . . . 28
  - 2.2.1 Metric Family . . . 28
  - 2.2.2 Correlation Family . . . 30
  - 2.2.3 Bregman Divergences . . . 32
  - 2.2.4 Distance Between Logical/Categorical Vectors . . . 35
  - 2.2.5 Mutual Information . . . 35
- 2.3 Divisive Clustering . . . 38
  - 2.3.1 $k$-means . . . 39
    - 2.3.1.1 Proof of Convergence . . . 40
  - 2.3.2 $k$-means++ . . . 41
  - 2.3.3 Fuzzy $c$-means . . . 42

2.3.3.1 Proof of convergence . . . 43
2.3.4 Soft $k$-means . . . 46
2.3.4.1 Proof of Convergence . . . 47
2.3.5 Gaussian Mixture Models . . . 50
2.3.5.1 Proof of Convergence . . . 52
2.4 Cluster Validity Indices . . . 58
2.4.1 Internal Validity Indices . . . 58
2.4.2 External Validity Indices . . . 63
2.5 Graph Theoretic Clustering . . . 64
2.5.1 Graph Preliminaries . . . 65
2.5.1.1 Graph Theoretic Definitions . . . 65
2.5.1.2 Constructing Graphs . . . 66
2.5.2 Agglomerative Linkage Clustering . . . 67
2.5.3 Maximum Flow: Edmonds-Karp-Dinitz algorithm . . . 69
2.5.3.1 Complexity Analysis . . . 73
2.5.4 Stoer-Wagner's Deterministic Minimum Cut Algorithm . . . 73
2.5.4.1 Complexity Analysis . . . 74
2.5.4.2 Proof of correctness . . . 75
2.5.5 Stochastic Flows . . . 76
2.5.5.1 Theoretical Framework . . . 77
2.5.5.2 Markov Clustering Algorithm . . . 79
2.5.5.3 Regularized Markov Clustering . . . 80
2.5.6 Spectral Clustering . . . 82
2.5.6.1 Ratio Cut . . . 83
2.5.6.2 Normalized Cut . . . 84
2.6 Consensus Clustering . . . 86
2.6.1 Definitions . . . 89
2.6.2 Monti's Consensus Clustering . . . 89
2.6.3 Evidence Accumulation . . . 92
2.6.4 Weighted Evidence Accumulation . . . 94
2.6.5 Fuzzy Evidence Accumulation . . . 94
2.6.6 Co-Association Tree . . . 95
2.7 Summary . . . 95

**3 Particle Swarm Optimization and Data Clustering 97**
3.1 Particle Swarm Optimization . . . 98
3.1.1 Convergence Analysis . . . 100
3.1.2 Complexity and Performance Analysis . . . 105
3.2 Van Der Merwe - Engelbrecht's PSO Clustering . . . 106
3.2.1 Performance Analysis . . . 107
3.3 Suboptimal Convergence in Higher Dimension . . . 107
3.3.1 Trajectory Analysis . . . 109
3.3.2 Analogy Using Game Theory . . . 111
3.3.3 Empirical Validation . . . 111
3.4 Summary . . . 113

**4 The Particle Swarm Clustering 115**
4.1 Definitions . . . 115
4.2 Algorithmic Framework . . . 118
4.2.1 Cohen - de Castro's Particle Swarm Clustering . . . 118

4.2.2 Szabo's Modified PSC (mPSC) . . . 118
4.2.3 Szabo's Fuzzy PSC (FPSC) . . . 120
4.3 Complexity Analysis . . . 120
4.3.1 Computational Complexity . . . 120
4.3.1.1 Computational Complexity of $\delta$ . . . 122
4.3.1.2 Computational Complexity of $\kappa$ . . . 124
4.3.2 Memory Complexity . . . 126
4.4 Trajectory Analysis . . . 128
4.4.1 Stability and Convergence . . . 128
4.4.2 Particle Behavior . . . 133
4.5 Performance Analysis . . . 134
4.5.1 Fisher - Iris Dataset . . . 136
4.5.1.1 General Overview . . . 136
4.5.1.2 Choosing the Distance Quantifier . . . 136
4.5.1.3 Cluster Validity Analysis . . . 137
4.5.2 Wine Dataset . . . 140
4.5.2.1 General Overview . . . 140
4.5.2.2 Choosing the Distance Quantifier . . . 142
4.5.2.3 Cluster Validity Analysis . . . 142
4.5.3 Spam E-mail Dataset . . . 144
4.5.3.1 General Overview . . . 144
4.5.3.2 Choosing the Distance Quantifier . . . 145
4.5.3.3 Cluster Validity Analysis . . . 146
4.6 Conclusion . . . 148

**5 Rapid Centroid Estimation 151**
5.1 The PSC and its Challenges . . . 151
5.1.1 Computational Complexity . . . 152
5.1.2 Memory Complexity . . . 153
5.1.3 Redundancies and Sensitivity to Initialization . . . 154
5.2 Definitions and Redefinitions . . . 156
5.3 Algorithmic Fundamentals . . . 162
5.4 Complexity Analysis . . . 162
5.4.1 Computational Complexity . . . 162
5.4.2 Memory Complexity . . . 164
5.4.3 Empirical Experiment . . . 165
5.5 Trajectory Analysis . . . 167
5.5.1 Stability and Convergence . . . 167
5.5.2 Particle Behavior . . . 170
5.6 Coping with Local Optima . . . 170
5.6.1 Substitution . . . 172
5.6.2 Particle Reset . . . 172
5.6.3 Swarm RCE: The Multi-Swarm Paradigm . . . 174
5.7 Experimental Results . . . 175
5.8 Conclusion . . . 178

**6 Ensemble Rapid Centroid Estimation 185**
6.1 Improving Swarm Diversity . . . 187
6.1.1 The Concept of "Charged Particles" . . . 187
6.1.2 Self-Evolution . . . 188

6.2 Ensemble Aggregation . . . . . . . . . . 189
6.2.1 Memory Complexity . . . . . . . . . . 190
6.2.2 Computational Complexity . . . . . . . . . . 192
6.3 Consensus Engrams . . . . . . . . . . 193
6.3.1 Definitions . . . . . . . . . . 193
6.3.2 Algorithmic Construct . . . . . . . . . . 197
6.3.3 Complexity Analysis . . . . . . . . . . 197
6.4 Applications . . . . . . . . . . 200
6.4.1 Color Image Segmentation . . . . . . . . . . 200
6.4.2 Vessel Extraction from Retinal Fundus Images . . . . . . . . . . 206
6.5 Conclusion . . . . . . . . . . 217

**7 Conclusion** **219**
7.1 Primary Findings . . . . . . . . . . 219
7.2 Limitations of the research . . . . . . . . . . 221
7.3 Recommendations for Future Work . . . . . . . . . . 222

**Bibliography** **223**

# List of Figures


1.1 % publications indexed in PubMed from 1990 to 2013 on the specific keywords/topics: Clustering, Data Mining, and Knowledge Discovery. . . . . . . . . . . . 6
1.2 Clustering a Magnetic Resonance Image. . . . . . . . . . . . . . . . . . . . . . . 7
1.3 Reconstructing HS images using RCE and Artificial Centroid Injection. . . . . 8
1.4 Clustering various retinal fundus images using ERCE. . . . . . . . . . . . . . . . 9
1.5 Estimating gait parameters using ERCE [17, 55] (cont.). . . . . . . . . . . . . . 10
1.5 Estimating gait parameters using ERCE. . . . . . . . . . . . . . . . . . . . . . . 11
1.6 Trajectory of the *swarm*{6} RCE$^{r+}$ 2014 particles (randomly seeded initial positions) recorded after 30 iterations on the five Gaussian dataset. . . . . . . . . 13
1.7 Clustering Non-convex dataset using ERCE. . . . . . . . . . . . . . . . . . . . 14
1.8 The algorithmic flow of ERCE explained using an artificial dataset . . . . . . . 15
1.9 Various Segmentation results using ERCE (Consensus Engrams). . . . . . . . . . 16
1.10 Variability of $K$ due to $\xi$ in ERCE swarms on the Half Rings dataset. . . . . . 19
1.11 Memory complexity for clustering 2-dimensional random noise (vol = 1 Byte – 1 Megabytes, d = 2) using various algorithms. The global settings for all algorithms are as follow: the number of representatives ($K$) = 30; the number of trials/swarms ($n_m$) = 30.) . . . . . . . . . . . . . . . . . . . . . . . . . 20
1.12 The CPU time and memory required of various algorithms with respect to the number of pixels in an image. . . . . . . . . . . . . . . . . . . . . . . . . . . 21
1.13 The ROC curve of the proposed method vs other methods in the literature [80]. 22
1.14 Research contribution diagram. . . . . . . . . . . . . . . . . . . . . . . . . . . 23

2.1 Clusters differ in term of their shape, size, and density. . . . . . . . . . . . . . 26
2.2 A possible taxonomy/interaction map of clustering algorithms . . . . . . . . . 27
2.3 Pairwise Euclidean distance, Manhattan distance, and Chebyshev distance . . 29
2.4 Minkowski distance with various $p$. . . . . . . . . . . . . . . . . . . . . . . . . 30
2.5 Mahalanobis distance with various $\Sigma$. . . . . . . . . . . . . . . . . . . . . . . . 30
2.6 Correlation distance between binary patterns and their fourier transforms. . . . 31
2.7 Mutual Information $I(X;Y)$ between $X$ and $Y$. . . . . . . . . . . . . . . . . . . 36
2.8 Conceptual illustration of the normalized mutual information. . . . . . . . . . . 38
2.9 Suboptimal partitions returned by $k$-means due to poor initialization . . . . . . 41
2.10 Fuzzy $c$-means clustering result on the five Gaussians dataset ($C$=5) . . . . . . 44
2.11 The resemblance between fuzzy $c$-means and soft $k$-means. . . . . . . . . . . . 48
2.12 GMM results on the Five Gaussians dataset. . . . . . . . . . . . . . . . . . . . . 53
2.13 Agglomerative clustering using single linkage criteria. . . . . . . . . . . . . . . 68
2.14 Illustration of agglomerative clustering using various linkage criteria . . . . . . 70
2.15 A simple illustration of the Edmonds-Karp-Dinitz algorithm for finding the maximum flow / minimum cut of a graph. . . . . . . . . . . . . . . . . . . . . . . . . 72
2.16 A simple illustration of Stoer-Wagner's algorithm. . . . . . . . . . . . . . . . . . 76
2.17 Illustration of MCL and R-MCL for clustering the Five Gaussians dataset . . . 82

2.18 Illustration of spectral (normalized cut) clustering on the Five Gaussians dataset 87
2.19 Various clustering results on non-convex clusters . . . . . . . . . . . . . . . . . . 88
2.20 The consensus matrices and dendrogram of five Gaussians data and uniformly distributed data . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 90
2.21 Stability of the five Gaussians data under various K . . . . . . . . . . . . . . . 92
2.22 The consensus distribution and CDF of five Gaussians data and uniformly distributed data . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 93
2.23 Illustration of the $K$-lifetime criterion on the Half Rings dataset. . . . . . . . . 94
2.24 Illustration of the CA-tree. . . . . . . . . . . . . . . . . . . . . . . . . . . . . 96

3.1 Convergence of a particle on 1-dimensional optimization problem with various stability criteria. . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 105
3.2 Initialization and Result: Van Der Merwe - Engelbrecht's PSO Clustering on a 2-Dimensional Dataset. . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 108
3.3 Trajectory of the Van Der Merwe - Engelbrecht PSO clustering vs K-means. $\measuredangle(v(t+1), \Delta) \to \pi/2$ as $K, dim \to \infty$. . . . . . . . . . . . . . . . . . . . . . . . 110
3.4 The Performance of PSO Clustering compared to $k$-means and $\text{RCE}^{\text{r+}}$ . . . . . 113
3.5 Trajectory of Van Der Merwe - Engelbrecht's PSO Clustering on an 8-Dimensional Dataset (4 times duplication of the two dimensions). Red square indicates swarm global best. The swarm converges to a severely suboptimal position. . . . . . . 114

4.1 Time complexity of the distance computation $\delta$. . . . . . . . . . . . . . . . . . 124
4.2 Time complexity of position update $\kappa$ of PSC, mPSC, and FPSC with varying $K$. 126
4.3 Memory complexity of various algorithms when clustering double precision floating point numbers. . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 127
4.4 Trajectory of the PSC/mPSC/FPSC particles on an artificial dataset . . . . . 135
4.5 Trajectory of the PSC/mPSC/FPSC particles on the five Gaussian dataset with numerous random seeding shows PSC's sensitivity to initialization. . . . . . . . 135
4.6 Fisher-Iris dataset. . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 137
4.7 A Succesful Clustering of the Fisher-Iris dataset ($\mathfrak{V}_P = 0.97$, algorithm: $\text{RCE}^{\text{r+}}$). 138
4.8 Six features out of the 13 from the wine dataset. . . . . . . . . . . . . . . . . . 140
4.9 A Succesful Clustering of the Wine dataset using Jensen-Shannon distance ($\mathfrak{V}_P = 0.97$, algorithm: $\text{RCE}^{\text{r+}}$). . . . . . . . . . . . . . . . . . . . . . . . . . . . . 143
4.10 A Succesful Clustering of the Spam dataset using Symmetrical Kulback-Leibler Divergence ($\mathfrak{V}_P = 0.901$, algorithm: $\text{RCE}^{\text{r+}}$). Spam and non-spam characteristics are expressed in terms of cumulative $\beta$-distribution. . . . . . . . . . . . . 147

5.1 Time complexity of the distance computation of PSC vs K-means. . . . . . . . 153
5.2 Complexity and Purity of the Benchmarked Algorithms. . . . . . . . . . . . . . 155
5.3 Memory complexity of various algorithms when clustering double precision floating point numbers. . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 166
5.4 Benchmarking the iteration complexity and overall memory complexity of various algorithm using a 3-classes 80-dimensional Gaussian dataset. . . . . . . . . 167
5.5 Trajectory of the RCE 2014 particles using various parameters . . . . . . . . . 171
5.6 Trajectory of the RCE 2014 particles on the five Gaussian dataset with numerous random seeding shows RCE's sensitivity to initialization. . . . . . . . . . . . . 171
5.7 Trajectory of the $\text{RCE}^{\text{r+}}$ 2014 particles on the five Gaussian dataset with numerous random seeding using both *substitution* and *particle reset* strategy. . . 173
5.8 Trajectory of the $swarm\{6\}$ $\text{RCE}^{\text{r+}}$ 2014 particles on the five Gaussian dataset with numerous random seeding using both *substitution* and *particle reset* strategy. 175

5.9 P-values to three decimal places based on the iteration times given in Table 5.5. Grey boxes indicate $p \leq 0.05$. . . . 178
5.10 P-values to three decimal places based on the resulting CH indices on all benchmark datasets given in Table 5.5. Grey boxes indicate $p \leq 0.05$. . . . 180
5.11 Box-plots [164, 165] of purity values given in Table 5.5. . . . 181
5.12 Box-plots [164, 165] of CH indices given in Table 5.5. . . . 182
5.13 Box-plots [164, 165] of optimization times given in Table 5.5. . . . 183

6.1 The algorithmic flow of ERCE explained using an artificial dataset . . . . 186
6.2 The typical growth curve of the swarm with varying entropy threshold. . . . 189
6.3 Variability of $K$ due to $\xi$ in ERCE swarms on the Half Rings dataset. . . . 189
6.4 Clustering Non-convex dataset using ERCE with charged particles. . . . 191
6.5 Memory complexity for clustering 2-dimensional random noise (vol = 1 Byte – 1 Megabytes, d = 2) using various algorithms. The global settings for all algorithms are as follow: the number of representatives ($K$) = 30; the number of trials/swarms ($n_m$) = 30.) . . . . 191
6.6 Comparison of the graphs produced by Gaussian Neighborhood, ERCE (CA-tree+fWEAC), and ERCE (Consensus Engrams). . . . 198
6.7 Finding the natural grouping in the BIRCH [167] datasets using ERCE with consensus engrams. . . . 199
6.8 Color Image Segmentation using ERCE (Consensus Engrams Scheme) on an image from Lotus Hill Institute [79]. . . . 201
6.9 The scatter plots of the two horses images in Figure 6.8 and its consensus engrams overlay. . . . 202
6.10 Image segmentation results using ERCE and their consensus engrams overlay. . 203
6.11 Image segmentation results using ERCE and their consensus engrams overlay — continued. . . . 204
6.12 The CPU time and memory required of various algorithms with respect to the number of pixels in an image. . . . 205
6.13 Illustration of % contamination. . . . 208
6.14 Feature extraction from a retinal fundus image. . . . 210
6.15 Scatter plot of the extracted features with the Consensus Engrams overlay. . . 211
6.16 Fundus Images (1–20) from DRIVE [80] arranged left to right, top to bottom. 212
6.17 Fundus Images (21–40) from DRIVE [80] arranged left to right, top to bottom. 213
6.18 Fuzzy segmentation results (gray) + detection mask (black) (1–20), arranged left to right, top to bottom. . . . 214
6.19 Fuzzy segmentation results (gray) + detection mask (black) (21–40), arranged left to right, top to bottom. . . . 215
6.20 The ROC curve of the proposed method vs other methods in the literature [80]. 216

# List of Tables


2.1 Agglomerative clustering: agglomeration criteria . . . . . . . . . . . . . . . . . . 68

3.1 Performance comparison between randomly seeded PSO clustering and $k$-means++ 112

4.1 Memory complexity of the PSC families vs other clustering algorithms . . . . . 128
4.2 Iris Features: P-value Matrix for testing the hypothesis of no correlation against the alternative that there is a non-zero correlation. . . . . . . . . . . . . . . . . . 136
4.3 Performance of the PSC families relative to other clustering algorithms on Fisher-Iris dataset (distance quantifier: Squared Mahalanobis Distance) . . . . . . . . 139
4.4 Wine Features: P-value Matrix for testing the hypothesis of no correlation against the alternative that there is a non-zero correlation. . . . . . . . . . . . 141
4.5 Performance of the PSC families relative to other clustering algorithms on Wine dataset (distance quantifier: Jensen-Shannon distance) . . . . . . . . . . . . . . . 144
4.6 Spam dataset feature overview, as reported in the original dataset . . . . . . . 145
4.7 Performance of the benchmarked clustering algorithms on Spam E-mail dataset (distance quantifier: Symmetrical Kullback-Leibler Divergence) . . . . . . . . . 148

5.1 Comparison between time complexities (milliseconds) and median purities of various algorithms on the experiments in the previous chapter . . . . . . . . . . 153
5.2 Comparison between memory complexities (bytes) and median purities of various algorithms on the experiments in the previous chapter . . . . . . . . . . . . . . 154
5.3 Worst case computational complexity of the RCE families vs PSC . . . . . . . 164
5.4 Memory complexity of the RCE, PSC, and other clustering algorithms . . . . . 165
5.5 Performance against UCI Machine Learning datasets [84] as reported in [4]. . . 176
5.6 Percentage improvement in time taken per iteration relative to PSC . . . . . . 177

6.1 Pros and Cons of the Concept of Charged Particles . . . . . . . . . . . . . . . . 188
6.2 Fundus Image Clustering Results Against the Gold Standard (Observer 1) summarized after 20 trials . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 209
6.3 Fundus Image Clustering Results Summary on the DRIVE database, appended with the results in [52, 57] . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 216

# Symbols

## *Variables and Constants*

| | |
|---|---|
| $a, b, c, n, v, w, x, y, z$ | scalars, unless otherwise noted |
| $K, C$ | number of clusters |
| $N, \lvert\mathbb{Y}\rvert$ | number of data |
| $n_m, \lvert\mathbb{S}\rvert$ | number of swarms |
| $n_i, \lvert\Theta\rvert$ | number of particles |
| $i, j, k, l, m, n$ | indices |
| $d, dim$ | dimension of a feature vector, used interchangeably |
| $t$ | time, iteration |
| $T$ | period |
| $f$ | function, frequency |
| $\mathbf{p}, \mathbf{g}, \mathbf{v}, \mathbf{x}, \mathbf{y}, \mathbf{z}$ | vectors |
| $\mathbf{y}_j \in \mathbb{R}^{dim}$ | the coordinate of the $j^{\text{th}}$ observation vector |
| $\mathbf{x}_i, \mathbf{z}_i \in \mathbb{R}^{dim}$ | the coordinate of the $i^{\text{th}}$ voronoi cell |
| $\mathbf{v} \in \mathbb{R}^{dim}$ | velocity vector |
| $\mathbf{so}, \mathbf{sc}, \mathbf{co}, \mathbf{mi} \in \mathbb{R}^{dim}$ | self-organizing, social, cognitive, and local minimum vectors |
| $\mathbf{P}, \mathbf{G}, \mathbf{V}, \mathbf{X}, \mathbf{Y}, \mathbf{Z}$ | matrices |
| $\mathbf{X}^M$,$\mathbb{X}^M$ | local minimum memory matrix |
| $\mathbf{I}$ | identity matrix |
| $\mathbb{S}, \mathbb{C}, \mathbb{V}$ | sets |
| $\emptyset$ | empty set |
| $\mathbf{u}, \mathbf{U}, \mathcal{U} \in \{0, 1\}$ | fuzzy responsibility (membership) vector / matrix |
| $\mathsf{u}, \mathsf{U} \in [0, 1]$ | crisp responsibility (membership) vector / matrix |
| | |
| $\pi$ | 3.141592653589793..., unless otherwise noted |
| $\Pi$ | linear mapping kernel matrix |
| $\sigma, std(X)$ | standard deviation |
| $\sigma^2, var(X)$ | variance |
| $\Sigma, cov(X)$ | covariance Matrix |
| $\rho$ | Pearson's correlation |

| | |
|---|---|
| $\Lambda, \lambda$ | Lagrange multiplier, eigenvalues |
| $\lambda_{(l)}$ | the constant for the $l^{\text{th}}$ term |
| $\varepsilon, th$ | threshold |
| $\alpha, \beta, \gamma, \pi, \kappa, \delta, \tau$ | various parameters |
| $\theta$ | multivariate Gaussian, particle |
| $\Theta$ | multivariate Gaussian Mixtures, swarm, subswarm |
| $\mathfrak{V}$ | cluster validity index |
| $\mathcal{C}$ | consensus matrix |
| $\varphi$ | random number, random vector |
| $\Omega$ | search space |
| $\eta\%$ | percentage |
| $\omega$ | inertia weight / velocity memory |
| $\psi_{j(l)} \in \mathbb{R}^{dim}$ | $j^{\text{th}}$ memory vector |

## *Set Operations*

| | |
|---|---|
| $\sup(\mathbb{S})$ | supremum (least upper bound) of a set |
| $\inf(\mathbb{S})$ | infimum (greates lower bound) of a set |
| $\mathbb{R}$ | set of real numbers |
| $\mathbb{R}^{dim}$ | $dim$-dimensional vector of real numbers |
| $x \in \mathbb{X}$ | $x$ in $\mathbb{X}$ |
| $x \notin \mathbb{X}$ | $x$ not in $\mathbb{X}$ |
| $\|\mathbb{A}\|$ | cardinality of a set |
| $\mathbb{A} \cup \mathbb{B}$ | union of sets $\mathbb{A}$ and $\mathbb{B}$ |
| $\mathbb{A} \cap \mathbb{B}$ | intersection of sets $\mathbb{A}$ and $\mathbb{B}$ |
| $\forall \mathbf{x} \in \mathbb{X}$ | for each vector $\mathbf{x}$ in $\mathbb{X}$ |
| $\exists \mathbf{x} \in \mathbb{X}$ | there exists $\mathbf{x}$ in $\mathbb{X}$ |
| $\mathbb{C} = \{\mathbb{C}_1, \ldots, \mathbb{C}_K\}$ | the clustered set $\mathbb{C}$, consisting of $K$ clusters $\{\mathbb{C}_1, \ldots, \mathbb{C}_K\}$ |
| $\mathbb{C}_x$ | a voronoi region with coordinate of the voronoi cell defined in $x$ |
| $y \in \mathbb{C}_x$ | the observation $y$, crisply associated to $\mathbb{C}_x$. |
| $\mathbb{A}^c$ | set complement |
| $\mathbb{B} \backslash \mathbb{A}$ | relative complement of $\mathbb{A}$ in $\mathbb{B}$ such that $\mathbb{B} \backslash \mathbb{A} = \{x \in \mathbb{B} \mid x \notin \mathbb{A}\}$ |

## *Matrix Operations*

| | |
|---|---|
| $\mathbf{X}^T$ | matrix transpose |
| $\mathbf{XY}$ | matrix multiplication |
| $\|\mathbf{X}\|$ | matrix determinant |
| $adj(\mathbf{X})$ | matrix adjoint |
| $tr(\mathbf{X})$, $Tr(\mathbf{X})$ | matrix trace |
| $\mathbf{X}^{-1}$ | matrix inverse |

| | |
|---|---|
| $\mathbf{X}^p$ | matrix power |
| $\mathbf{X} \circ \mathbf{X}$ | Hadamard product (elementwise multiplication) |
| $\mathbf{U}^T * \mathbf{U}$ | Pairwise fuzzy T-norm operator |
| $diag(\mathbf{X})$ | the diagonal components of $\mathbf{X}$ |
| $diag(\mathbf{x})$ | diagonal matrix which diagonal values are $\mathbf{x}$ |

## *Probability Functions*

| | |
|---|---|
| $p(x), q(x)$ | probability distribution functions |
| $p(x\|z, \theta)$ | conditional probability of $x$ given $z$ and the model $\theta$ |
| $p(x\|z, \Theta)$ | conditional probability of $x$ given $z$ and the mixture model $\Theta$ |
| $p(x, y)$ | joint probability of $x$ and $y$ |
| $E[X]$ | expectation of $X$ |
| $E[X\|Y]$ | conditional expectation of $X$ given $Y$ |
| $\mathcal{L}(\theta)$ | likelihood function |
| $H(X)$ | Shannon's entropy of $X$ |
| $H(X\|Y)$ | conditional entropy of $X$ given $Y$ |
| $H(X, Y)$ | joint entropy of $X$ and $Y$ |
| $I(X; Y)$ | mutual information between $X$ and $Y$ |
| $NMI(X, Y)$ | normalized mutual information between $X$ and $Y$ |
| $KL(P\|\|Q), D(P\|\|Q)$ | Kullback-Leibler Divergence between $P$ and $Q$ |
| $D_{JSD}(P\|\|Q)$ | Jensen-Shannon Divergence between $P$ and $Q$ |
| $\mathcal{P}(x)$ | empirical probability distribution |

## *General Functions*

| | |
|---|---|
| $\mu \pm \sigma$ | average plus and minus standard deviation |
| $x > y$ | $x$ greater than $y$ |
| $x \geq y$ | $x$ greater than or equal to $y$ |
| $x < y$ | $x$ less than $y$ |
| $x \leq y$ | $x$ less than or equal to $y$ |
| $x \to \infty$ | $x$ approaches infinity |
| $x \leftarrow x + 2y$ | assign $x + 2y$ as a new value for $x$ |
| $\langle \mathbf{x}, \mathbf{y} \rangle$ | inner product of $\mathbf{x}$ and $\mathbf{y}$ |
| $\partial f(x, y)/\partial x$ | partial derivative of $f(x, y)$ with respect to $x$ |
| $\|x\|$ | absolute value of $x$ such that $\|x\| = x$ if $x \geq 0$ and $\|x\| = -x$ if $x < 0$ |
| $\nabla_{\mathbf{z}} f(\mathbf{z})$ | gradient of $f(\mathbf{z})$ |
| $\log(x)$ | natural logarithm of x, $\ln(x)$ |
| $\exp(x)$ | exponent of x, $e^x$ |
| $\sum_{x \in \mathbb{X}} f(x)$ | sum of $f(x)$ for all $x$ in $\mathbb{X}$ |

| | |
|---|---|
| $\frac{1}{\lvert\mathbb{X}\rvert}\sum_{x\in\mathbb{X}} f(x)$ | average of $f(x)$ for all $x$ in $\mathbb{X}$ |
| $\frac{\sum_i w_i f(x_i)}{\sum_i w_i}$ | weighted average of $f(x)$, indexed with $i$ |
| $\prod_{x\in\mathbb{X}} f(x)$ | product of $f(x)$ for all $x$ in $\mathbb{X}$ |
| $\phi(x)$ | kernel function |
| $\mathcal{O}(N)$ | complexity function |
| $y = \arg\max_x f(x)$ | $y$ is the $x$ that maximizes $f(x)$ |
| $y = \arg\min_x f(x)$ | $y$ is the $x$ that minimizes $f(x)$ |
| $d(\mathbf{x}, \mathbf{y}), d_{xy}$ | distance between $\mathbf{x}$ and $\mathbf{y}$ |
| $D(\mathbf{X}, \mathbf{Y})$ | pairwise distance between vectors in $\mathbf{X}$ and $\mathbf{Y}$ |
| $\Psi(\mathbf{x})$ | resultant vector function for the voronoi cell $\mathbf{x}$ |

*To Mom and Dad,*

*To Nina*

# List of Publications

1. Mitchell Yuwono, Steven W. Su, Bruce D. Moulton, and Hung T. Nguyen. An algorithm for scalable clustering: Ensemble rapid centroid estimation. *Conf Proc of the IEEE Congress on Evolutionary Computation*, pages 1–8, 2014.

2. Mitchell Yuwono, Steven W. Su, Ying Guo, Bruce D. Moulton, and Hung T. Nguyen. Unsupervised nonparametric method for gait analysis using a waist-worn inertial sensor. *Applied Soft Computing*, 14, Part A: 72 – 80, 2014. ISSN 1568-4946. Special issue on hybrid intelligent methods for health technologies.

3. Mitchell Yuwono, Steven W. Su, Bruce D. Moulton, and Hung T. Nguyen. Data clustering using variants of rapid centroid estimation. *IEEE Transactions on Evolutionary Computation*, 18: 3: 366–377, 2013.

4. Mitchell Yuwono, Steven W. Su, Ying Guo, Jiaming Li, Sam West and Josh Wall. Automatic Feature Selection Using Multiobjective Cluster Optimization for Fault Detection in a Heating Ventilation and Air Conditioning System. *Conf Proc IEEE International Conference on Artificial Intelligence, Modelling, and Simulation*, 2013: 1–6, 2013.

5. Mitchell Yuwono, Steven W. Su, Bruce D. Moulton, and Hung T. Nguyen. Unsupervised segmentation of heel-strike IMU data using rapid cluster estimation of wavelet features. *Conf Proc IEEE Eng Med Biol Soc*, 2013: 953–956, 2013.

6. Mitchell Yuwono, Steven W Su, and Bruce D. Moulton. An Approach to Fall Detection using Gaussian Distribution of Clustered Knowledge. *Bio-Informatic Systems, Processing and Applications*. River Publishers. 2013: 69 – 82, 2013.

7. Mitchell Yuwono, Bruce D. Moulton, Steven W. Su, Branko G. Celler, and Hung T. Nguyen. Unsupervised machine-learning method for improving the performance of ambulatory fall-detection systems. *Biomedical Engineering Online*, 11: 9, 2012.

8. Mitchell Yuwono, Steven W. Su, Bruce D. Moulton, and Hung T. Nguyen. Gait episode identification based on wavelet feature clustering of spectrogram images. *Conf Proc IEEE Eng Med Biol Soc*, 2012: 2949–2952, 2012.

9. Mitchell Yuwono, Jonathan Sepulveda, and A. M. Ardi Handojoseno. Centroid extraction from Hartmann-Shack images using swarm clustering approach. *Conf Proc IEEE Eng Med Biol Soc*, 2012: 1446–1449, 2012.

10. Mitchell Yuwono, Steven W. Su, Bruce D. Moulton, and Hung T. Nguyen. Gait cycle spectrogram analysis using a torso-attached inertial sensor. *Conf Proc IEEE Eng Med Biol Soc*, 2012: 6539–6542, 2012.

11. Mitchell Yuwono, Steven W. Su, Bruce D. Moulton, and Hung T. Nguyen. Fast unsupervised learning method for rapid estimation of cluster centroids. In *Proc. of the 2012 IEEE Congress on Evolutionary Computation*, pages 889–896, June 10–15 2012.

12. Mitchell Yuwono, Steven W. Su, Bruce D. Moulton, and Hung T. Nguyen. Optimization strategies for rapid centroid estimation. In *Proc. of the 34rd Annual International Conference of the IEEE EMBS*, pages 6212–6215, San Diego, Aug. 28–sept. 1 2012.

13. Mitchell Yuwono, Steven W. Su, Bruce D. Moulton, and Hung T. Nguyen. Method for increasing the computation speed of an unsupervised learning approach for data clustering. In *Proc. of the 2012 IEEE Congress on Evolutionary Computation*, pages 2957–2964, June 10–15 2012.

14. Mitchell Yuwono, A. M. Handojoseno, and Hung T. Nguyen. Optimization of head movement recognition using Augmented Radial Basis Function Neural Network. *Conf Proc IEEE Eng Med Biol Soc*, 2011: 2776–2779, 2011.

15. Mitchell Yuwono. Unwrapping hartmann-shack images of off-axis aberration using artificial centroid injection method. In *IEEE Conference in Biomedical Engineering and Informatics (BMEI)*, pages 560–564, 2011.

16. Mitchell Yuwono, Steven W. Su, and Bruce D. Moulton. Fall detection using a Gaussian distribution of clustered knowledge, augmented radial basis neural-network, and multilayer perceptron. *Conf Proc 6th International Conference on Broadband and Biomedical Communications (IB2Com)*, 2011: 145–150, 21-24 Nov. 2011. doi: 10.1109/IB2Com.2011.6217909