Developing a Cohesive Space-Time Information Framework for Analyzing Movement Trajectories in Real and Simulated Environments

by

Atsushi Nara

A Dissertation Presented in Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy

Approved July 2011 by the
Graduate Supervisory Committee:

Paul Torrens, Chair
Soe Myint
Michael Kuby
William Griffin

ARIZONA STATE UNIVERSITY

December 2011

ABSTRACT

In today's world, unprecedented amounts of data of individual mobile objects have become more available due to advances in location aware technologies and services. Studying the spatio-temporal patterns, processes, and behavior of mobile objects is an important issue for extracting useful information and knowledge about mobile phenomena. Potential applications across a wide range of fields include urban and transportation planning, Location-Based Services, and logistics. This research is designed to contribute to the existing state-of-the-art in tracking and modeling mobile objects, specifically targeting three challenges in investigating spatio-temporal patterns and processes; 1) a lack of space-time analysis tools; 2) a lack of studies about empirical data analysis and context awareness of mobile objects; and 3) a lack of studies about how to evaluate and test agent-based models of complex mobile phenomena. Three studies are proposed to investigate these challenges; the first study develops an integrated data analysis toolkit for exploration of spatio-temporal patterns and processes of mobile objects; the second study investigates two movement behaviors, 1) theoretical random walks and 2) human movements in urban space collected by GPS; and, the third study contributes to the research challenge of evaluating the form and fit of Agent-Based Models of human movement in urban space. The main contribution of this work is the conceptualization and implementation of a Geographic Knowledge Discovery approach for extracting high-level knowledge from low-level datasets about mobile objects. This allows

better understanding of space-time patterns and processes of mobile objects by revealing their complex movement behaviors, interactions, and collective behaviors. In detail, this research proposes a novel analytical framework that integrates time geography, trajectory data mining, and 3D volume visualization. In addition, a toolkit that utilizes the framework is developed and used for investigating theoretical and empirical datasets about mobile objects. The results showed that the framework and the toolkit demonstrate a great capability to identify and visualize clusters of various movement behaviors in space and time.

ACKNOWLEDGMENTS

I would never have been able to complete my dissertation without the help, support, guidance, and efforts of many people.

Foremost, I would like to gratefully acknowledge Dr. Paul Torrens for his excellent guidance, caring, patience, and friendship during my graduate studies. His mentorship and support helped me overcome many situations and finish this dissertation. I would also like to thank my committee members, Dr. William Griffin, Dr. Michael Kuby, and Dr. Soe Myint for insightful comments and constructive suggestions at different stages of my research.

Many thanks go to Dr. Kiyoshi Izumi as well as members of the Digital Human Research Center at the National Institute of Advanced Industrial Science and Technology, the Department of Systems Innovation at the University of Tokyo, and Institute of Advanced Biomedical Engineering and Science at the Tokyo Women's Medical University for providing me a research opportunity and wonderful time while I was in Tokyo, Japan.

I thank my colleagues of the School of Geographical Sciences and Urban Planning at the Arizona State University. In particular, I would like to thank members of the Geosimulation Research Laboratory, Haojie Zhu, Xun Li, and Scott Brown for their helps, discussions, and friendships. A very special thank you to my friend, Dr. Darren Ruddell for his help and friendship.

The last, and most importantly, I want to deeply thank my family for their love and unconditional supports. I thank my wife for her love, encouragement,

sacrifice and being with me through the best and worst moments. I also thank my son, Jinta, for inspiring and amazing me every day. I thank my parents, Ryoichi and Akiko, for their support and faith in me. Also I thank Mayumi's parents, Hitoshi and Motoko for their support and encouragement.

TABLE OF CONTENTS

Page

LIST OF TABLES x

LIST OF FIGURES xii

CHAPTER

1 INTRODUCTION 1

2 LITERATURE REVIEW 8

2.1 Classic Geographical Approaches 9

2.2 Behavioral Geography 12

2.2.1 Time Geography 15

2.2.2 Travel Choice Behavior and Activity-based Approach 19

2.2.3 Route Choice, Navigation, and Orientation 21

2.2.4 Collective Behavior and Pedestrian Crowds 24

2.3 Location Aware Technology 28

2.3.1 Location Estimation Methods 29

2.3.1.1 Triangulation 30

2.3.1.2 Proximity 33

2.3.1.3 Scene analysis 34

2.3.2 Location Aware Systems 34

2.3.2.1 Outdoor environments 35

2.3.2.2 Indoor environments 37

2.4 Geographic Knowledge Discovery 39

2.4.1 Knowledge Discovery from Databases 41

CHAPTER Page

2.4.2 Geographic Knowledge Discovery 42

2.4.3 Trajectory Data Mining 45

2.5 Complex Systems and Agent-Based Models 48

3 RESEARCH OBJECTIVES 56

4 SPATIO-TEMPORAL ANALYSIS AND VISUALIZATION OF MOBILE OBJECTS 60

4.1 Overview 60

4.2 Related Works 61

4.3 Methodology 65

4.3.1 Quantitative Analysis of Mobile Objects 66

4.3.1.1 Velocity and acceleration 66

4.3.1.2 Sinuosity 68

4.3.1.3 Fractal dimension 69

4.3.1.4 Power-law distribution 71

4.3.1.5 Directional statistics 72

4.3.2 Qualitative Visualization of Mobile Objects based on Time Geography 74

4.3.2.1 Visualization of space time path 74

4.3.2.2 Volume rendering 84

4.3.3 Space-Time Analysis Toolkit 85

4.4 Case Study - Pedestrian Evacuation Dynamics 87

4.4.1 Dataset 88

CHAPTER Page

4.4.2 Results of Quantitative and Qualitative Analysis of Crowd Evacuation Dynamics 91

4.5 Discussion and Conclusions 100

5 TRAJECTORY DATA MINING: CLUSTERING, CONTEXT RECOGNITION, AND SPATIO-TEMPORAL VISUALIZATION 103

5.1 Overview 103

5.2 Related Works 105

5.3 Methodology 112

5.3.1 Trajectory Partitioning 114

5.3.2 Trajectory Clustering 122

5.3.3 Evaluation of Trajectory Clustering 126

5.3.3.1 Behavioral recognition by decision tree 126

5.3.3.2 Visualization of trajectory cluster distribution 127

5.3.3.3 Trajectory data mining tool 129

5.4 Results 130

5.4.1 Trajectory Data Mining on Simulated Data 131

5.4.1.1 Dataset 131

5.4.1.2 Results 135

5.4.2 Trajectory Data-Mining on GPS Data 169

5.4.2.1 Dataset 169

5.4.2.2 Results 172

CHAPTER Page

5.5 Discussion and Conclusions 203

6 EVALUATION OF A PEDESTRIAN SIMULATION MODEL BY TRAJECTORY DATA MINING APPROACH 207

6.1 Overview 207

6.2 Related Works 208

6.2.1 Pedestrian Movement and Evacuation Behaviors 209

6.2.2 Modeling Pedestrian Dynamics and Evacuation Behaviors 212

6.2.3 Model Evaluation 217

6.2.4 Research Objectives 221

6.3 Methodology 222

6.3.1 Pedestrian Evacuation Simulation based on Social Force Model 223

6.3.2 Trajectory Data-Mining for Evaluating Pedestrian Dynamics in Agent-Based Model 225

6.4 Results 232

6.4.1 Simulation Scenarios and Dataset 232

6.4.2 Evaluating Simulation Scenarios 235

6.4.2.1 Descriptive statistics of trajectories 235

6.4.2.2 Trajectory data-mining: clustering and visualization 241

6.4.2.2.1 Global analysis of behavioral cluster 248

CHAPTER Page

6.4.2.2.2 Temporal analysis 250

6.4.2.2.3 Spatio-temporal analysis 258

6.5 Discussion and Conclusion 289

7 SUMMARY 293

7.1 Achievements and Findings 295

7.1.1 Study 1 295

7.1.2 Study 2 297

7.1.3 Study 3 298

7.2 Limitations 299

7.3 Discussions and Future Works 301

REFERENCES 305

LIST OF TABLES

Table Page

1. Comparison among LATs 35
2. Data-mining tasks and techniques 42
3. Motion descriptors of trajectories 92
4. Correlation matrix of trajectories' motion descriptors 93
5. Number of sub-trajectories in each partitioning algorithm 137
6. Correlation matrix for movement variables (Sim: no partition)..... 142
7. Correlation matrix for movement variables (Sim: TRACLUS-MDL) 142
8. Correlation matrix for movement variables (Sim: Distance-Threshold) 143
9. Results of PCA (Sim: No Partition) 143
10. Results of PCA (Sim: TRACLUS-MDL) 144
11. Results of PCA (Sim: Distance-Threshold) 145
12. Results of decision tree classification 164
13. Confusion matrix of behavioral recognition 164
14. Frequency of activities 171
15. Correlation matrix for movement variables (GPS: no partition).... 176
16. Correlation matrix for movement variables (GPS: TRACLUS-MDL) 177
17. Correlation matrix for movement variables (GPS: Distance-Threshold) 177

Table Page

18. Results of PCA (GPS: no partition) 178

19. Results of PCA (GPS: TRACLUS-MDL)..... 179

20. Results of PCA (GPS: Distance-Threshold) 180

21. Results of decision tree (Main activity) 197

22. Result of decision tree (Binary activity)..... 197

23. Behavioral match between clusters and real activities..... 198

24. Descriptive statistics of trajectory's motion descriptors 239

25. Descriptive statistics of segment's motion descriptors 240

LIST OF FIGURES

Figure Page

1. Space-Time Path and Space-Time Prism 16
2. Location-sensing technologies: Location accuracy and scale of deployment 29
3. Lateration technique to estimate two dimensional location 31
4. Angulation technique to estimate two dimensional location..... 33
5. Velocity of a mobile object 67
6. Radar plots. Direction change frequency distribution of Porcupine Caribou Herd (PCH) sample (Left) and direction change frequency distribution of caribou individual Blixen (right) 72
7. STP (left: single path, right: multiple paths colored by path ID). 76
8. STPs colored by velocity value (left: single path, right: multiple paths)..... 76
9. Stream tubed STP (left: single path, Inverse velocity, right: acceleration)..... 76
10. Stream tubed STPs (left: multiple paths, Inverse velocity, right: acceleration)..... 77
11. Image of assigning motion descriptors to vertices in a STP. 82
12. Partitioning a space-time aquarium and averaging motion descriptors. 83
13. Intersection between a segment and a plane. 83
14. GUI of main display of the Space-Time Analysis toolkit. 86

Figure Page

15. GUI of the Space-Time Kernel Density Estimation Tool 87

16. Simulation snapshots of pedestrian evacuation dynamics (time unit: frame). 90

17. Trajectories of pedestrian evacuation dynamics. 91

18. A 2D map of trajectories (left: evacuation time, right: average velocity). 93

19. A 2D map of trajectories (left: path length, right: straight length)... 94

20. A 2D map of trajectories (left: straightness index, right: fractal dimension)..... 94

21. A 2D map of trajectories (circular dispersion). 94

22. Histogram: A frequency distribution of length of segments 95

23. A 2D map of trajectories 95

24. A 2D KDE map of trajectories 96

25. Stream-tubed STPs colored by average velocity of segments 98

26. Stream-tubed STPs colored by average acceleration of segments ... 98

27. Space-Time volume density maps 99

28. Trajectory partition (TRACLUS with the MDL approach). 116

29. Three components of the distance function in TRACLUS 117

30. Formation of the MDL cost 119

31. Trajectory partition (Distance-Threshold approach). 121

32. Labeling staying behavior on sub-trajectories..... 121

33. Mapping temporal cluster distribution. 128

Figure Page

34. GUI of the trajectory data mining tool (partitioning & clustering). 130

35. Trajectories of BM 133

36. Trajectories of CRW 134

37. Trajectories of Lévy flight 134

38. Mixed trajectories of BM, CRW, and Lévy flight 135

39. Trajectory partitioning results of Brownian motion. 138

40. Trajectory partitioning results of Correlated Random Walk. 138

41. Trajectory partitioning results of Lévy Flight. 138

42. Trajectory partitioning results of Lévy Flight in Space-Time Cube. 139

43. Gap curve for three partitioning algorithms (Simulation). 146

44. Number of subtrajectories in a cluster (n=300) (no partition: *k*=3). 147

45. Number of subtrajectories in a cluster (n=300) (no partition: *k*=13). 147

46. Number of subtrajectories in a cluster (n =46,441) (TRACLUS-MDL: *k*=2). 148

47. Number of subtrajectories in a cluster (n =46,441) (TRACLUS-MDL: *k*=4). 148

48. Number of subtrajectories in a cluster (n =40,335) (Distance-Threshold: *k*=5). 149

Figure Page

49. Number of subtrajectories in a cluster (n =40,335) (Distance-Threshold: *k*=19). 149

50. Cluster profile (no-partition: *k*=3). 152

51. Cluster profile (no-partition: *k*=13). 152

52. Cluster profile (TRACLUS-MDL: *k*=2)..... 153

53. Cluster profile (TRACLUS-MDL: *k*=4)..... 153

54. Cluster profile (Distance-Threshold: *k*=5). 154

55. Cluster profile (Distance-Threshold: *k*=19). 154

56. Trajectory clusters (no partition: *k*=3). 155

57. Trajectory clusters (no partition: *k*=13). 155

58. Sub-trajectory clusters (TRACLUS-MDL: *k*=2)..... 156

59. Sub-trajectory clusters (TRACLUS-MDL: *k*=4)..... 156

60. Sub-trajectory clusters (Distance-Threshold: *k*=5)..... 156

61. Sub-trajectory clusters (Distance-Threshold: *k*=19)..... 157

62. Temporal cluster distribution (no partition). 159

63. Temporal cluster distribution (TRACLUS-MDL). 159

64. Temporal cluster distribution (Distance-Threshold). 160

65. Five samples of misclassified trajectories in CRW using the no partitioning approach. 160

66. A tree visualization of Decision Tree results (Distance-Threshold: *k*=5). 165

67. STPs of BM colored by cluster ID. 165

Figure Page

68. STPs of CRW colored by cluster ID. 166

69. STPs of Lévy Flight colored by cluster ID. 166

70. Space-Time line density map of Cluster 1. 167

71. Space-Time line density map of Cluster 2. 167

72. Space-Time line density map of Cluster 3. 168

73. Space-Time line density map of Cluster 4. 168

74. Space-Time line density map of Cluster 5. 169

75. Study area and GPS trajectories. 172

76. GPS trajectories in the area of a subject's residence. 172

77. Two-tone STP representation of trajectory partitioning (TRACLUS-MDL). 174

78. Two-tone STP representation of trajectory partitioning (Distance Threshold). 174

79. Gap curve for three partitioning algorithms (GPS). 180

80. Number of subtrajectories in a cluster (n=36) (no partition: k=5). 182

81. Number of subtrajectories in a cluster (n=36) (TRACULS-MDL: k=3). 182

82. Number of subtrajectories in a cluster (n=36) (TRACULS-MDL: k=5). 183

83. Number of subtrajectories in a cluster (n=36) (Distance-Threshold: k=4). 183

Figure Page

84. Number of subtrajectories in a cluster (n=36) (Distance-Threshold: *k*=8). 184

85. Cluster profile (no partition: *k*=5). 187

86. Cluster profile (TRACLUS-MDL: *k*=3)..... 187

87. Cluster profile (TRACLUS-MDL: *k*=5)..... 188

88. Cluster profile (Distance-Threshold: *k*=4). 188

89. Cluster profile (Distance-Threshold: *k*=8). 189

90. Sub-trajectory clusters (no partition: *k*=5)..... 189

91. Sub-trajectory clusters (TRACLUS-MDL: *k*=3)..... 190

92. Sub-trajectory clusters (TRACLUS-MDL: *k*=5)..... 190

93. Sub-trajectory clusters (Distance-Threshold: *k*=4)..... 190

94. Sub-trajectory clusters (Distance-Threshold: *k*=8)..... 191

95. Temporal cluster distribution. 194

96. A tree visualization of Decision Tree results (Major activity: Distance-Threshold, *k*=8)..... 201

97. A tree visualization of Decision Tree results (Binary activity: Distance-Threshold, *k*=8)..... 201

98. Space-Time line density map in morning activity (Distance-Threshold, *k*=8). 202

99. Schematic overview of evaluation procedures for ABMs of mobile objects. 226

Figure Page

100. Three-dimensional map algebra using local function and arithmetic operator of subtraction. 231

101. Street designs for four simulation scenarios (numbers represented are in simulation unit length). 234

102. Trajectories of four simulation scenarios. 235

103. STPs colored by average velocity (unit length/unit time) of segment (Pedestrian evacuation) 240

104. The gap curve for identifying the optimal *k* value. 242

105. Trajectory clustering framework and result. 246

106. Cluster profiles for pedestrian evacuation simulation. 247

107. Proportion of clusters in each scenario. 250

108. Temporal cluster distribution of evacuees in each scenario. 254

109. Summarized temporal cluster distribution in Scenario 1. 255

110. Summarized temporal cluster distribution in Scenario 2. 255

111. Summarized temporal cluster distribution in Scenario 3. 256

112. Summarized temporal cluster distribution in Scenario 4. 256

113. Comparison of summarized temporal cluster distribution between Scenario 1 and 2. 257

114. Comparison of summarized temporal cluster distribution between Scenario 1 and 3. 257

115. Comparison of summarized temporal cluster distribution between Scenario 1 and 4. 258

116. STPs of movements with moderate velocity described by Cluster 1. Color of path represents average segment velocity of a sub-trajectory (unit length / second). 261

117. STPs of fragmented paths described by Cluster 2. Color of path represents average segment velocity of a sub-trajectory (unit length / second). 262

118. STPs of successful evacuation described by Cluster 3. Color of path represents average segment velocity of a sub-trajectory (unit length / second). 263

119. STPs of successful evacuation described by Cluster 4. Color of path represents average segment velocity of a sub-trajectory (unit length / second). 264

120. STPs of slow movement described by Cluster 5. Color of path represents average segment velocity of a sub-trajectory (unit length / second). 265

121. STPs of slow movement described by Cluster 6. Color of path represents average segment velocity of a sub-trajectory (unit length / second). 266

122. STPs of successful evacuees from the West corridor described by Cluster 7. Color of path represents average segment velocity of a sub-trajectory (unit length / second). 267

Figure Page

123. STPs of clogging behavior described by Cluster 8. Color of path represents average segment velocity of a sub-trajectory (unit length / second). 268
124. Space-Time line density map (Cluster 1: moderate velocity). 269
125. Space-Time line density map (Cluster 2: fragmented path). 270
126. Space-Time line density map (Cluster 3: smooth & continuous). . 271
127. Space-Time line density map (Cluster 4: smooth & continuous). . 272
128. Space-Time line density map (Cluster 5: clogging). 273
129. Space-Time line density map (Cluster 6: clogging). 274
130. Space-Time line density map (Cluster 7: smooth & continuous). . 275
131. Space-Time line density map (Cluster 8: clogging). 276
132. Difference of cluster density distribution between scenarios (Cluster 1). 280
133. Difference of cluster density distribution between scenarios (Cluster 2). 281
134. Difference of cluster density distribution between scenarios (Cluster 3). 282
135. Difference of cluster density distribution between scenarios (Cluster 4). 283
136. Difference of cluster density distribution between scenarios (Cluster 5). 284

Figure Page

137. Difference of cluster density distribution between scenarios (Cluster 6). 285

138. Difference of cluster density distribution between scenarios (Cluster 7). 286

139. Difference of cluster density distribution between scenarios (Cluster 8). 287

140. Detail difference of cluster density distribution between Scenario 1 and 2. 288

Chapter 1

INTRODUCTION

In today's world, an unprecedented amount of data about individual mobile objects, from micro to macro scales in space and time, have become more available. This is particularly due to the development and deployment of location aware technologies (LATs), the emergence of ubiquitous computing environments, and the usefulness of these techniques in everyday life. While collecting data about mobile objects with LATs might be limited by cost, privacy, and security issues, an alternative data source of mobile objects is a simulation model, which offers great capability in generating massive amounts of realistic details of spatio-temporal movement. Simulated data are particularly useful for situations which are difficult to identify or to test through real-world observation or experiments with LATs. When considered together, data from the real-world and simulation offer opportunities for investigating spatio-temporal patterns and behaviors of mobile objects in completely new ways.

Studying the spatio-temporal patterns, processes, and behaviors of mobile objects is an important and current research task, as extraction of useful information and knowledge about dynamic and mobile phenomena is driven by real demands in various applications; for example, vehicle and pedestrian traffic control for transportation management and facilities design, Location-Based Services (LBS) (e.g., navigation assistance and mobile advertising); weather forecasting (e.g., hurricane trajectory prediction and risk analysis); law enforcement (e.g., video surveillance for criminal activities); animal conservation

(e.g., tracking at-risk animal populations); and logistics management for goods and human. As supplies of rich, complex, and ubiquitous data of mobile objects grow, and demands from various real-world applications increase, spatio-temporal analysis and modeling of mobile objects has become a major challenge for the scientific community, across domains from Geographic Information Science (GISci), computer science, and engineering, to biology, and social and behavioral science. Understanding behaviors of massive mobile objects is a challenge largely due to the behavioral complexity. For example, pedestrian dynamics consist of complex movement behaviors at multi scales such as macro-scale trip planning, meso-scale route choice and way finding, and micro-scale locomotion. In addition, such behaviors are not only affected by personal factors such as preference, experience, and knowledge but also environmental factors such as route structure, available transportation, and situations along the route. Moreover, non-linear interactions among other individuals as well as interactions between individual and environment introduce further complexity with feedback, scaling effects, and path dependence.

The research that I propose is designed to contribute to the existing state-of-the-art in tracking and modeling mobile objects, in particular targeting challenges in investigating spatio-temporal patterns and processes by making use of the unprecedented individual-based data now available. Specifically, this research focuses on the following challenges; 1) a lack of space-time analysis tools; 2) a lack of studies about empirical data analysis and context awareness (semantics) of movement datasets, particularly those considered as trajectories;

and 3) a lack of studies about how to evaluate and test Agent-Based Models (ABMs) of mobile phenomena particularly focusing on a complex spatio-temporal and behavioral process of mobile agents.

First, there is an increasing demand on effective and efficient tools to extract hidden patterns, trends, and useful information and knowledge from spatio-temporal datasets, which are often unprecedentedly massive, high-dimensional, and complex (e.g., heterogeneous data sources, multivariate connections, explicit and implicit spatial and temporal relations and interactions) (Mennis & Guo, 2009). Miller (2003) mentioned the importance of developing spatio-temporal data mining and exploratory visualization techniques to handle very large, detailed, and noisy space-time attribute data. Peuquet (2002) also discussed that GIS (Geographic Information Systems) and GIS users need to filter through vast amounts of data to find patterns and associations in addition to traditional GIS tasks of database manipulation, analysis, and visualization. Over the last two decades, many efforts have been made in studying space-time patterns and processes, in particular implementing the concept of Hägerstrand's time geography, often in a GIS environment (e.g., Kwan and Hong, 1998; Kwan, 1998(a); Kapler and Wright, 2004; Miller, 1991; Miller and Han, 2001; Shaw et al., 2008; Shaw and Yu, 2009; Yu, 2006). Despite the fact that these efforts have demonstrated the strong capability of GIS to represent and analyze individual activities in a space-time context (Shaw, Yu, & Bombom, 2008), research challenges exist in furthering quantitative and qualitative investigations and

related tool developments, so that hidden patterns and trends in the complex individual-based spatio-temporal data of mobile objects can be more explored.

Second, in recent years there has been increasing interest in studying movements by trajectory-based data mining that can infer patterns from new sets of massive amounts of data that are passively and automatically generated. In trajectory data mining, data of individual mobile objects are considered as sequences of the location and timestamp of a mobile object. Using a set of spatio-temporal sequences of mobile objects, trajectory data mining discovers spatio-temporal knowledge through data mining exercises including pattern detection, clustering, classification, generalization, outlier detection, and visualization. There are considerable research examples that propose trajectory data mining algorithms and methodologies; however, most of them have focused on the geometric shape of trajectories without taking into account the context of the data (Bogorny, Kuijpers, & Alvares, 2009). In addition, the few trajectory data mining methods that have been implemented and applied in practice (Dodge, Weibel, & Forootan, 2009), are being developed in a rather piecemeal fashion, and have yet to migrate from research to demonstrate convincing social and commercial benefits (Weibel, Sack, Sester, & Bitterlich, 2008). Thus, further exploration and investigation are required to advance the development of theory, methodology, and practice for the extraction of useful information and knowledge from massive and complex trajectory databases.

Third, ABM is a useful approach for modeling movement behaviors with several benefits such as capturing emergence and modeling flexibility. In ABM, a

system is modeled as a collection of autonomous agents, which possess characteristics of heterogeneous, proactive, perceptive, communicative, and adaptive. Local interactions of such individual agents can describe surprising patterns of emergent phenomena, for example, pedestrian lane formation as a self-organizing phenomenon. In addition, ABM is flexible to model system environments as well as agent behaviors, which is particularly useful for spatial simulations (Smith, Goodchild, & Longley, 2009). For example, ABM can define various types of system environment such as continuous space, road networks, and building as well as agent's attributes and behaviors such as preferences, perception of neighborhoods, and movement modes. Thus, ABM can be used as a tool for exploring and experimenting with existing theories and ideas as an artificial laboratory with high degrees of realism and detail. A key research challenge in ABM is model evaluation, to examine how well simulated results represent real behaviors of mobile objects. Model evaluation is a general term for model calibration, verification, and validation. Respectively, calibration, verification, and validation involve: 1) specifying or fitting a model (fine tuning the model to some dataset); 2) ensuring that it functions and it is internally consistent (testing the logic of model structure, e.g. seeing if models work in different software platforms and with different data); 3) and comparing model structure and outcomes with information not used in model construction (measuring the goodness of fit). Particularly model validation is a difficult task when systems in the real-world as well as generated by ABM exhibit complex behaviors, such as feedback, path-dependence, phase shift, non-linearity,

emergence, adaptation, and self-organization. Which aspects of the model behavior are to be compared with empirical data is a research challenge. Complex behaviors cannot be simply examined by looking at global statistics, but it is necessary to consider spatio-temporal process and behaviors across various scales. Developing an analytical framework for model comparison to empirical data is also useful to compare simulation outcomes from what-if scenarios.

This dissertation research aims to investigate all three research challenges—*in a cohesive and interconnected approach*—by conducting three studies on spatio-temporal analysis and modeling of human movement. The first study develops an integrated spatio-temporal data exploration tool to represent spatio-temporal patterns and process of mobile objects and seeks to contribute to the first research challenge (i.e., methods and tools for extracting trajectory data from large and complex spatio-temporal datasets). The second study offers insight into the research challenge of space-time data analysis by focusing on generating and associating context to trajectories. Applying the tool developed in the first study and extending it by adding a trajectory data mining method, it explores the spatio-temporal pattern and process of two movement datasets; 1) theoretical random walks and 2) human movements in urban space collected by Global Positioning System (GPS). The third study contributes to the research challenge of evaluating dynamic (computer) models of human behavior in urban space, by applying the developed tool in the first and second studies to quantitatively and qualitatively evaluate the form and fit of a computer model of movement under what-if scenarios.

The overarching goal of this research is to improve upon the current state-of-the-art in spatio-temporal analysis and modeling of complex human movement. To achieve this goal, I conducted three cohesive and interconnected studies on human trajectory data based around tool development, space-time analysis, visualization, data mining, simulation, and model evaluation. Three major achievements of this dissertation include; the usefulness of the developed toolkit to quantitatively and qualitatively investigate spatio-temporal pattern and process of mobile objects; the extraction of complex behavior and knowledge about mobile objects that are hidden under trajectory datasets; and the usefulness of the trajectory data mining tool for extracting collective movement behaviors and evaluating ABMs.

The rest of the dissertation is organized as follows. A literature review is presented in Chapter 2, describing the current state-of-the-art in space-time analysis, semantic data analysis, and spatial modeling. The literature review sets that context for Chapter 3, in which I outline my research objectives and in which I discuss the novelty that my research will contribute to the existing state-of-the-art. Chapters 4, 5, and 6 of the dissertation are devoted to describing the approach that I have developed in addressing these objectives, with more specific details of the methodology, research design, and results of empirical analysis to be deployed. Ultimately, these will form three distinct and independent research sub-projects, each of which is interconnected. Chapter 7 of the dissertation provides summary and concluding remarks.

Chapter 2

LITERATURE REVIEW

Space and time are inseparable components of reality. People, animals, goods, information, and many entities in our world move over space and time; all the while, they commonly leave location and trajectory traces (often in digital form) (Laube, 2005). Focusing on human movement, human geographers have long been studying human spatio-temporal patterns and processes across different scales in space and time; for example, international/interregional migration at a macro-scale (Mark & Wright, 2005); intra-urban household relocation and daily trips in a city at a meso-scale (Clark & Huang, 2003); and pedestrian movements on a street as a micro-scale (Batty, 2003). Detailed and heterogeneous individual behaviors, dynamic processes, and complex interactions of individuals and their environment at multi spatio-temporal scales are usually important in explaining and understanding such geographical phenomena because different behaviors and influences manifest at different scales and the connections between them are complex. This chapter reviews theoretical backgrounds and relevant studies about the behavior, analysis, and modeling of mobile objects. The chapter is organized as follows. Section 2.1 discusses the limitations in traditional spatial analysis. Section 2.2 presents approaches of behavioral geography including sub-sections of time geography, decision making and choice behavior of residential mobility, navigation and orientation, and collective behavior and pedestrian crowd phenomena. Section 2.3 explains location aware technologies. Section 2.4 describes knowledge discovery from spatial databases and briefly introduces

trajectory data mining approaches. Section 2.5 reviews issues in complex system and ABM.

2.1 Classic Geographical Approaches

Classic approaches in geography such as conventional location theories commonly look at geographical phenomena in a way that is relatively coarse, aggregate, static, normative, and inflexible across scales (Batty, 2005). Scholarship in these topics has often adopted a reductionist view, with the result that traditional approaches have several limitations of representing geographical phenomena in the real-world, particularly when they are embedded as the theoretical foundations for models and analysis (Batty, 2005).

First, classic approaches of spatial models are relatively weak in handling spatial detail. Therefore, there is often a disparity between models and reality on a behavioral level. In particular, many models adopt a reductionist view of systems, i.e., a top-down approach. (Regional science is an example of this.) A reductionist approach is one that addresses complexity in a system by decomposing the system into constituent components and gaining an understanding of their interactions in the process. In some cases, this approach works well, particularly in situations where the whole system can be pieced together from a sum of smaller parts. However, when processes that operate at the local level are interdependent, the reductionist approach faces the challenges of the ecological fallacy (Wrigley, Holt, Steel, & Tranmer, 1996) and modifiable areal unit problems (MAUP) (Openshaw, 1983). These problems occur when an inference about individual level attributes

or behaviors is drawn from data about aggregates so that an understanding of the processes that generate macro-scale patterns may not be easily developed by simply aggregating up from the individual. In addition to the coarse representation of reality, traditional approaches are often spatially inflexible, meaning that a model represents a phenomenon at one scale. It is, however, important for spatial models to accommodate a wide variety of spatial scales, ideally in an integrated and seamless manner that is capable of generating realistic behaviors that can be considered across many levels of observation. For example, approaches in regional science such as the input-output model (Isard, Azis, Drennen, Miller, Saltzmann, & Thorbecke, 1998) deal with macroscopic analysis based on aggregated information; therefore, they cannot infer any disaggregate behaviors reliably.

Second, geographical models should be capable of capturing the ability of phenomena to evolve over time because many geographical phenomena are dynamic. Traditional spatial models represent time as static, and usually with poor temporal resolution. Some models use cross-sectional data, which are collected for a single period in time, or a snapshot, while others use longitudinal data that are a series of snapshots often separated by long periods of time (Torrens, 2002, p. 210). Thus, these models constitute a weak proxy for dynamics. For instance, McHugh and Gober (1992) studied the interstate migration system using annual state-to-state migration flow data from Internal Revenue Service (IRS) and their findings demonstrated the higher degree of temporal and spatial volatility in the

U.S. interstate migration system as compared to traditional migration studies using the decennial census population dataset.

Third, traditional spatial models often lack representations of behavioral process. For instance, spatial interaction models estimate the volume of flows between origin and destination based on structural attributes of two areas (e.g., population density, employment opportunity, floor space) (Wilson, 1975; Fotheringham & O'Kelly, 1989). However, even when a model can estimate or predict the flows and movements of goods, people, and information over networks by connecting hierarchically arranged nodes with accuracy, there is usually little explanation in the model of *how* and *why* those movements occurred from individual behaviors.

Fourth, traditional geographical theories and models are often based on abstract assumptions. In many classic urban models (Thünen, 1826; Alonso, 1960; Fujita, 1982), the geographic variability of landscape is assumed as a uniform plain; transportation is assumed to be available equally in all directions at a similar cost; people are assumed to have the same utilities and preferences for good, services, and products; populations are assumed to be constant, not expanding, and to consist of uniform ethnic or cultural memberships; and decision making and choice behavior is assumed to be economically and spatially rational, in which human has perfect knowledge and the ability to make optimal decisions that maximize utility (Golledge R. G., 2008). In spatial theories, the maximization of utility has usually been assumed to result from the minimization of transportation costs or, in the simplified case, from the minimization of physical

distance (De la Barra, 1989). Nevertheless, human behaviors in reality cannot be always described by such neat, abstracts assumptions. Studies of household trips and expenditures, for example, often show substantial difference between the distances that household members travel to make the nearest and maximum purchases of goods as opposed to where conventional theories (e.g., central place theory), expected them to go (Golledge, Rushton, & Clark, 1966; Golledge R. G., 2008). Except for a few consumer activities that are classified as convenience goods and services (e.g., grocery purchases, attendance at church, and gasoline purchases), many other goods and services are typified by shopping-around activities, which could not be described by a least effort/least cost/least distance syndrome (Golledge R. G., 2008).

2.2 Behavioral Geography

During the late 1960s and early 1970s, researchers in human geography argued that classic approaches in geography such as conventional location theories were not satisfactory to describe geographical phenomena; in particular, they were weak in explaining the understanding of the relationship between the dynamics of human behavior and the dynamics of the environment, i.e., human-environment interaction. Behavioral geographers, therefore, replaced simplistic and mechanistic conceptions of human-environment relations with a new perspective that explicitly recognized the complexities of human behavior (Walmsley & Lewis, 1984).

Essential ingredients of behavioral geography, as set out by Golledge and Timmermans (1990, p. 57) are:

- "*A search for models of humanity which were alternatives to the economically and spatially rational beings of normative location theory;*
- *A search to define environments other than objective physical reality as the milieu in which human decision making and action took place;*
- *An emphasis on processural rather than structural explanations of human activity and relationship between human activity and the physical environment;*
- *An interest in unpacking the spatial dimensions of psychological, social, and other theories of human decision-making and behavior;*
- *A change in emphasis from aggregate populations to the disaggregate scale of individuals and small groups;*
- *A need to develop new data sources other than the generalized mass-produced aggregate statistics of government agencies which obscured and overgeneralized decision making processes and consequent behavior;*
- *A search for methods other than those of traditional mathematics and inferential statistics that could aid in uncovering latent structure in data, and which could handle data sets that were less powerful than the traditionally used interval and ratio data;*
- *A desire to merge geographic research into the ever-broadening stream of crossdisciplinary investigation into theory building and problem solving*."

With these perspectives, approaches in behavioral geography aim at understanding of human-environment interaction by looking at both the psycho-socio-spatial processes of individual cognition about the (social, physical) environment and the way in which these processes influence the nature of resultant behavior (Walmsley & Lewis, 1984). In particular, the behavioral approach emphasizes human decision-making and choice behavior in the context of the role of spatial cognition, which deals with spatial knowledge, knowing, intelligence, and reasoning by humans. Spatial cognitive structures and processes include those of sensation, perception, thinking, learning, memory, attention, imagination, bias, conceptualization, language, and reasoning and problem solving of spatial properties including location, size, distance, direction, separation and connection, shape, pattern, and movement (Montello, 2001; Montello, 2009). In this chapter, four research topics of behavioral geography in relation to pedestrian movement are particularly highlighted, because of variations in approach with variations in the scale of observation of behavior. These topics include: time geography; decision-making and choice behavior for activity scheduling (i.e., macro-scale movement and behavioral geography); routing choice, navigation, and wayfinding behavior (i.e., meso-scale movement and behavioral geography); and one-to-one and one-to-many interactions in collective movement and crowd behavior (i.e., micro-scale movement and behavioral geography). There exist opportunities to capture some of these aspects of behavioral geography using LATs, data, and next-generation GIS.

2.2.1 Time Geography

Space-time activities of individuals have increasingly become the focus of research by behavioral geographers and GIScientists particularly due to the technological advancements in LATs that allow for tracking of dynamics of mobile objects such as animals, vehicles, and humans. Geographers see new opportunities to study behaviors of mobile objects and have called for reconsideration of a conceptual framework of Hägerstrand's time geography (Hedley, Drew, Arfin, & Lee, 1999). Because time and space play an inseparable role in human activities, Hägerstrand proposed the concept of time geography to study the relationship between human activities and various constraints in a space-time context (Golledge & Stimson, 1997). In its theoretical framework, individual's activities are limited by three constraints; 1) capability constraints are the physical and technological limitations such as sleeping and auto ownership respectively; 2) coupling constraints are anchors on activity that enable people to bundle their activities to places and times (work, home, school, etc.); and 3) authority constraints are temporal and/or spatial limitations or regulations on space-time accessibility, as in the case of military areas (spatial constraints) and office hours (temporal constraints) (Yu & Shaw, 2007).

With these constraints controlling the spatio-temporal patterns of individual activities, the two fundamental concepts/constructs of time geography—space-time path and space-time prism—were proposed to illustrate spatio-temporal characteristics of human activities. A space-time path, known as

STP or space-time lifeline, is an individual's trajectory in space and time, which begins and has an origin at the point of birth and ends and has a destination at the point of death. It is usually represented visually on a two-dimensional (really, 2.5D) plane that shows geographical positions (x,y axis) and uses a perpendicular dimension (z axis) to represent time (Figure 1, Left). A STP provides an event-oriented framework for analyzing individual's activities based on spatial and temporal change with space and time constraints (Hägerstrand, 1970; Lenntorp, 1976). A space-time prism describes the extent in space and time that an individual can access under a specific set of constraints (Figure 1, Right).

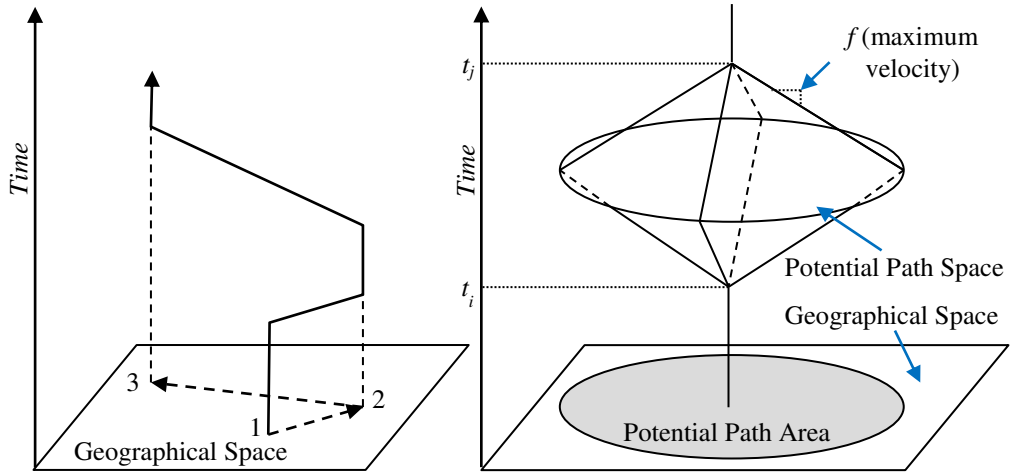

Figure 1. Space-Time Path and Space-Time Prism. Adapted from Miller (2005).

There are wide a variety of explorations and applications of time geography. In developing a framework for visualizing time geography, Andrienko, Andrienko, & Gatalsky (2003) introduced the cube perspective of exploratory spatio-temporal visualization: a space-time cube is used to encapsulate the volume

of space and time occupied by activities. Kraak & Koussoulakou (2004) developed a visualization environment for the space-time cube. Oculus Info, Ltd. introduced an integrated GIS/STP visualization and database environment, *Geotime*, as a tool for displaying and tracking individual-based events, objects, and activities in space and time within a single, interactive 3-D view (Kapler & Wright, 2004). Miller (1991; 1999) applied the principle of the space-time cube in attempting to establish accessibility measures in an urban environment. As an extension of this, Miller and Bridwell (2009) recently formulated analytical definitions of the STP and prism, in which unobserved components are characterized by minimum cost curves through an inverse velocity field in order to capture complex velocities.

In terms of analytical techniques, Corcoran, Higgs, Brunsdon, & Ware (2007) applied three techniques to investigate spatio-temporal patterns of property fires and vehicle fires. They are; 1) temporal analysis, in which simple line and circular plots were used for different granularities of time; 2) spatial analysis, in which the fire incident concentration was explored through the use of a cumulative sum technique based on wards, and a kernel density method was applied to highlight spatial variability and to show how these variations changed by incident category; and 3) spatio-temporal analysis, in which the technique of comap was used to illustrate spatio-temporal dynamics. Laube, Dennis, Forer, & Walker (2007) focused on quantifying individual motion data. The authors proposed a dynamic perspective to analysis, which referred to the variability of motion properties throughout four lifeline context operators; 1) instantaneous; 2)

interval; 3) episodal; and 4) total. Yu (2006) developed analytical functions that identify four different spatio-temporal patterns among people through their STPs; co-existence, co-location in space, co-location in time, and no co-location in either space or time.

As applied studies, Kwan (1998; 1999) used time geography to study urban accessibility differences by gender and different ethnic groups. Moore, Whigham, Holt, Aldridge, & Hodge (2003) developed the SCRUM (Spatio-Chronological Rugby Union Model) to recode and visualize a rugby game under a time geography framework. Space-Time analysis is not only limited to human activities but also any kind of mobile objects can be studied using the approach. Kritzler, Raubal, & Krüger (2007) automatically tracked movements of laboratory mice based on a passive RFID (radio-frequency identity tag) sensor system and developed a module for visualization and analysis. Ware, Arsenault, Plumlee, & Wiley (2006) visualized spatio-temporal behaviors of whales and revealed several behavioral patterns in their tracks.

The framework of time geography captures spatio-temporal complexity on mobile objects; however, it has been used mainly as a conceptual model, partially due to the limitation the lack of a computational environment to implement the framework effectively (Yuan, Mark, Egenhofer, & Peuquet, 2004). Recent technological advancements in computational environment now allow us to handle very large spatio-temporal dataset of mobile objects along with the huge data influx of mobile objects collected by LATs and generated by ABMs. Simultaneously, database researchers have extended conventional spatial

databases and designed spatio-temporal databases for handling and querying mobile object data in databases (Wolfson, 2002; Goerge & Shekhar, 2006).

As described above, many efforts have been made in studying space-time patterns and processes of human activities and interactions, often in a GIS environment. While those efforts demonstrated that GIS can provide a powerful platform to represent and analyze individual activities in a space-time context (Shaw, Yu, & Bombom, 2008), research challenges exist in furthering quantitative and qualitative investigations for collective movements and related tool developments, so that hidden patterns and trends in the complex individual-based spatio-temporal data of multiple mobile objects can be more explored.

2.2.2 Travel Choice Behavior and Activity-based Approach

As a macro-scale movement behavior of individuals (e.g., pedestrians, households), this section discusses behavioral geography and the activity-based approach.

Human lives consist of activities such as working, socializing, shopping and recreation, activities of which require time and space that are often available at particular locations for limited durations (Miller, 2004). The location and timing of such key activities differ by individuals and depend on available time, transportation, and communication resources to conduct these activities (Miller, 2004). Human travel behaviors within cities can be described by the activity-based approach, which has been an active research topic over the past few decades particularly in transportation research (Ettema & Timmermans, Theories

and models of activity-travel patterns, 1997). The fundamental concept of the activity-based approach is that travel is derived from the participation in activities instead of being pursued for its own sake; therefore, the understanding, analysis, and forecasting of travel behavior should be based on the understanding of activities (Burnett & Hanson, 1982; Joh, Arentze, & Timmermans, 2001). Individuals try to meet their personal and family needs by participating in activities in everyday life (e.g., work, shopping, and recreation), subject to a set of constraints including space, time budget, physical environment, and various individual-oriented factors such as socio-economic characteristics, cultural environment, and personal preference. Thus, travel behavior is derived as a by-product to overcome the distance between activity locations in the process of organizing activities in time and space weighted by various constraints (Joh, Arentze, & Timmermans, 2001).

Modeling of disaggregated travel behavior based on the activity-based approach has been dominated by the use of the discrete choice model (DCM), the origin of which is in models of consumer choice, microeconomic theory, psychological judgment theory, and statistical analysis of categorical data. In DCMs, it is assumed that people allocate time according to the principle of utility maximization, i.e., "within the time and cost constraints imposed by their budgets, people choose to spend time in activities which is proportional to their (process and/or goal achievement) utilities" (Axhausen & Gärling, 1992, p. 326). The activity-based approaches utilizing DCMs have been provided the understandings of, for example, the characteristics of trip chaining (Damm & Lerman, 1981;

Kitamura, Nishii, & Goulias, 1990), choice of activity participation and duration (Kitamura, 1984), choice of activity patterns (Adler & Ben-Akiva, 1979; Recker, McNally, & Roth, 1986), and the structure of activity pattern explained by spatial, temporal and interpersonal constraints (Pas & Koppelman, 1987; Pas, 1988). However, Gärling (1994; 1998) pointed out that while the utility maximization principle might explain which factors affect the final choice, it does not account for the process of making decisions that also impact on outcomes. This problem is not an issue if a research objective is to estimate travel demand; nevertheless, it is an important factor for a better understanding of travel behavior. In addition, many activity-based models have failed to account for the highly dynamic nature of activity participation, i.e., continuous decision-making process (Ettema, 1996). These problems are critical factors for understanding pedestrian behaviors because their decision-makings and choice behaviors (e.g., path planning, way finding, avoid collisions, and find attractions) are influenced by interactions with other pedestrians as well as their surrounding environment that dynamically change; thus, their decision-makings and choice behaviors also need to be dynamically updated.

2.2.3 Route Choice, Navigation, and Orientation

While decision-making and choice behavior of activity scheduling discussed in the previous section focuses at a scale of strategic level in a city (macro-scale), route choice, navigation, and orientation are human movement behavior at a scale of tactical level on a street (meso to micro scale).

Empirical studies revealed characteristics of pedestrian route choice behaviors such as subconscious and directness (Hill, 1982), and preference to follow the shortest route as primary strategy (Ciolek, 1981; Senevarante & Morall, 1986; Gärling & Gärling, 1988). Other factors that are considered to affect pedestrian choice behaviors include personal factors such as age, gender, preferences (Bovy & Stern, 1990), past experience (Golledge & Stimson, 1997), and trip characteristics such as trip purpose (e.g., sightseeing, work-related walking trip) (Bovy & Stern, 1990), route structures (e.g., sidewalks, paved, tree), and situations along the route (e.g., traffic volume, attractive spots).

Similar to modeling of activity scheduling, route choice behaviors have been modeled by DCMs. Such models are based on the theoretical assumption that all actions of the pedestrian, let it be performing an activity or walking along a certain route, will provide utility (i.e., induce cost) to him and he will predict and optimize this expected utility, taking into account the uncertainty in the expected traffic conditions (Hoogendoorn & Bovy, 2004). DCMs are analytically tractable and could be calibrated with real-world data from activity surveys and travel diaries (Torrens, 2011).

Applying DCMs, Gipps (1986) described pedestrian route choice behaviors through the walking facility by determining a finite number of routes through the walking infrastructure. While human routing choice behavior has been often studied with network-based models, which are suitable for vehicle applications because movements of vehicles are unidirectional and limited by discrete number of decision points (nodes). Therefore vehicle travelers choose a

route from a limited number of route alternatives. Contradictory to this, the number of pedestrian route alternatives should not be restricted because of pedestrian's freedom of movement in public space; therefore, the network-based approaches are generally less applicable (Hoogendoorn & Bovy, 2004). By relaxing the discrete network assumption, Hoogendoorn and Bovy (2004) developed a dynamic mixed discrete-continuous choice approach to modeling pedestrian route and activity choice behavior in public facilities, in which route alternatives are continuous functions in time and space.

Navigation involves the behavioral process of one's movement from origin of one's location toward pre-selected destination along the pre-defined route (Golledge R. G., 2004). Understanding navigation processes, behavior, and cognitive aspects of accessibility is important for not only theoretical investigations in spatial cognition but also practical applications such as developing navigation systems for travelers, visually-impaired persons, and even autonomous robots. According to Montello (2005), there are two components of navigation: locomotion and wayfinding; "locomotion is the movement of one's body around an environment, coordinated specifically to the local or proximal surrounds – the environment that is directly accessible to our sensory and motor systems at a given moment, [whereas] wayfinding is the goal-directed and planned movement of one's body around an environment in an efficient way" (Montello, 2005, p. 259). Locomotion behaviors, the process of which takes place in the vicinity of a person's local surroundings, include behaviors such as

identifying open spaces ahead, steering to avoid obstacles/collisions, and finding and organizing movement relative to landmarks (Montello, 2005).

Wayfinding behaviors deal with navigation at a large extent and thus involve managing, planning, and deciding about trip routes, waypoints, and the chaining/scheduling of trips in particular sequences or frequencies (Montello, 2005). Both are related to orientation, which "refers to a person's ability to relate personal location to environmental frames of reference" (Golledge & Stimson, 1997, p. 511). Sadalla and Montello (1989) discussed that there are two orientation reference systems; 1) allocentric orientation, using external features such as landmarks or coordinate systems (e.g., north-south, east-west); and 2) eccentric orientation, which depends on one's body position (e.g., left/right, in front/behind). For instance, while in the former system a person may update one's orientation based on visible landmarks, in the later system one may use dead reckoning updating that involves updating by inferring a new location/heading based on knowledge about movement speed and direction from a known starting point, without recognition of specific features (Montello, 2009).

2.2.4 Collective Behavior and Pedestrian Crowds

Collective behavior deals with the interrelated and connected activity of people in groups, often with a similar or coordinated response to events or stimuli. This can include people who all occupy the same location (e.g., a street crowd and riots), as well as mass phenomena in which individuals are dispersed across a wide area (e.g., social movements and trends) (Forsyth, 2009).

Collective behaviors have been studied theoretically for a long time and many collective behaviors have been identified, some of which are geographical in nature. One example is contagion theory, developed by a social psychologist, Gustave Le Bon (1895). His theory suggests that behavior (especially emotional behavior) can be passed/transmitted between people in the same way that germs pass through contagion and this can explain why groups sometimes behave in the same way (Forsyth, 2009). In this theory, "the anonymity of the crowd, along with other conditions, results in the loss of individual rationality, leaving crowd members especially susceptible to suggestions from others in the crowd and to common emotional and destructive impulses. Because of this, crowd behavior is volatile and spontaneous (Schweingruber & Wohlstein, 2005, p. 144)." However, many researchers who observed collective behaviors (e.g., riot) claimed that there are discrepancies between the contagion theory and empirical observations. First-hand observations by, for example, Turner and Killian (1987) and McPhail (1994) revealed that individual behavior within a crowd is neither as anonymity nor as irrational as the contagion theorists believed.

Convergence is another theoretical explanation for group behavior. This is different than the contagion hypothesis: "convergence theory assumes that individuals who join rallies, riots, movements, crusades, and the like all possess particular personal characteristics that influence their group-seeking tendencies. Such aggregations are not haphazard gatherings of dissimilar strangers; rather, they represent the convergence of people with compatible needs, desires, motivations, and emotions" (Forsyth, 2009, p. 516). In other words, people

assemble into groups (perhaps in the same space and time) because of shared goals or intent; this is different than the "averaging" hypothesis suggested by contagion theory.

In terms of spatial consideration of crowd behavior, McPhail & Tucker (1992) studied individual and collective actions in temporary gatherings based on perception control theory. Perception control theory, developed by Powers (1973), argues that each separate individual is trying to control his or her experience in order to maintain a particular relationship to others, i.e., a spatial relationship with others in a group. In their study, a simulation system, GATHERING, was developed and graphically shows movement, milling, and structural emergence in crowds (see the discussion by Thalmann and Musse (2007)). The same simulation system was later used by Schweingruber (1995) to study the effects of reference signals common to coordination of collective behavior and by Tucker, Schweingruber, & McPhail (1999) to study formation of arcs and rings in temporary gatherings.

Several researchers studied collective behaviors in terms of self-organizing phenomena. Self-organization means that the patterns of collective behaviors are not externally planned, prescribed, or organized, for example, by traffic signs, laws, or behavioral conventions (Helbing & Molnár, 1997), but the spatio-temporal patterns emerge through the nonlinear interactions of individuals. The organization in self-organization is therefore usually spatial or spatio-temporal. For example, Reynolds (1987) built a model to simulate the motion of a flock of birds, named boids. There are originally three rules that explain the boid

movement behaviors; (1) Avoidance, (2) Copy, and (3) Center, and later the fourth rule, View, was added by Flake (2001). Avoidance is to move away from boids that are too close so that reducing the chance of collisions. Copy is to fly in the general direction that the flock is moving by averaging the other boids' velocities and directions. Center is to minimize exposure to the flock's exterior by moving toward the perceived center of the flock. View is to move laterally away from any boid that blocks the view. These simple rules are applied to each individual locally, yet interactions with other individuals result the complex nature of flocking behavior (Flake, 2001).

Helbing (1992), Helbing, Keltsch, & Molnár (1997), and Helbing, Farkas, & Vicsek (2000) proposed a model based on physics and sociopsychological forces to describe collective behaviors of pedestrians. The model was set up as a particle system and the change of velocity with time *t* is given by the dynamic equation as follows.

$$m_i \frac{dv_i}{dt} = m_i \frac{v_i^0(t)e_i^0(t) - v_i(t)}{\tau_i} + \sum_{j(\neq i)} f_{ij} + \sum_w f_{iw}$$

where, each particle *i* of mass m_i had a predefined speed v^0_i, i.e., the desired velocity, in a certain direction e^0_i to which it tends to adapt its instantaneous velocity v_i within a certain time interval τ. Simultaneously, the particles try to keep a velocity-dependent distance from other entities *j* and wall *w* controlled by interaction forces f_{ij} and f_{iw}, respectively.

Without assuming strategical considerations, communication, or imitative behavior of pedestrians (Helbing & Molnár, 1997), the model (according to which

individuals behave rather automatically) can explain the self-organized pedestrian collective behavior of crowds; the formation of lanes consisting of pedestrians with the same desired walking direction; oscillatory changes of the walking direction at narrow passages; and the temporary emergence of unstable roundabout traffic with an alternating rotation direction at intersections (Helbing & Molnár, 1997). The social force model, however, has little bearing on theory because movements of pedestrians are treated as purely based on physics without intelligence and with largely homogenous characteristics and behaviors, although Daamen and Hoogendoorn (2003) have run real-world experiments and the model works for some examples.

2.3 Location Aware Technology

While the previous section reviews theoretical views of movement behaviors in geography, this section discusses how to collect real data about behavioral geography and individual movement, how to the data with massive volumes of objects, and how to automate the data collection process via location-aware technologies (LATs).

In recent years, various types of LATs have been developed. Figure 2 describes the general relationship between location accuracy and scale of deployment of LATs; each box's horizontal span shows the range of accuracies the technology covers; the bottom boundary of each box represents current deployment; and the top boundary shows predicted deployment in the near future (Hazas, Scott, & Krumm, 2004). These LATs differ with respect to the location

estimation methodology as well as specifications of devices such as accuracy, coverage, frequency of location updates, and cost of installation and maintenance.

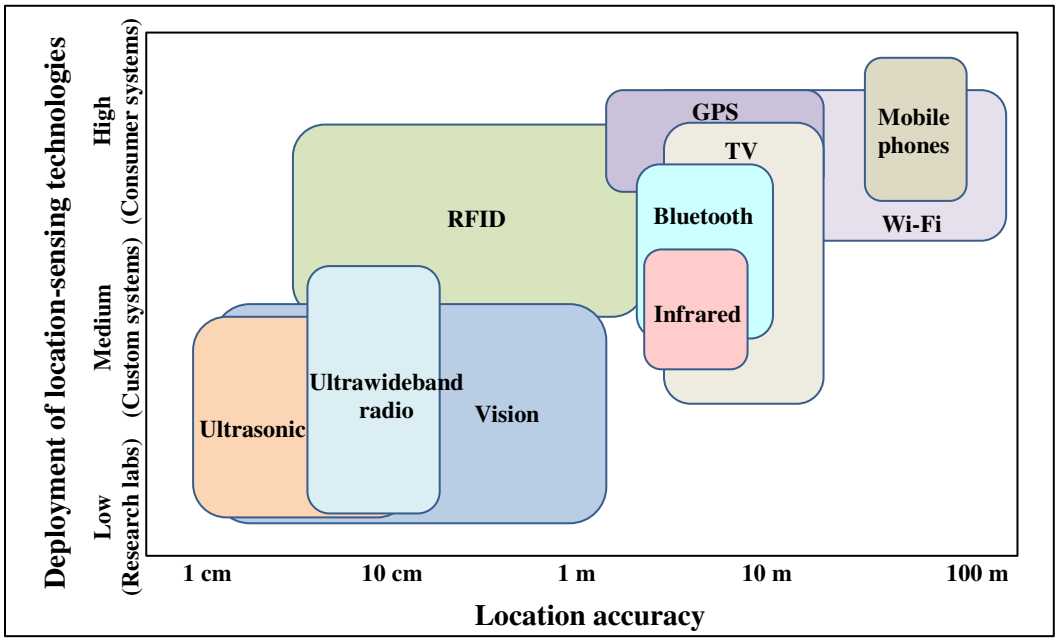

Figure 2. Location-sensing technologies: Location accuracy and scale of deployment. Adapted from Hazas, Scott, & Krumm (2004).

2.3.1 Location Estimation Methods

Location is at the core of understanding movement and geographic behavior because movement behavior is described by locational changes and the behavior is strongly tied with existing geographical contents (e.g., what are the physical, economical, social, and cultural environments at a certain location and its neighbors?). Thus, accurately and automatically determining objects' locations is desirable.

Location estimation is to determine an object's location with respect to a reference point. There are three principal techniques for location estimation; triangulation, proximity, and scene analysis, and they can be employed in a location system individually or in combination (Hightower & Borriello, 2001). For example, assisted GPS (Global Positioning System) combines proximity-based location sensing for increasing speed to obtain satellite signals with triangulation-based GPS for better location estimation.

2.3.1.1 Triangulation

The triangulation technique uses the geometric properties of triangles to calculate object locations by cross-referencing their geometry. Two common types of triangulation technique include lateration, which relies on distance measurement, and angulation, which relies on angle measurement (Hightower & Borriello, 2001).

Lateration: It estimates the position of an object by measuring its distance from multiple reference positions (Hightower & Borriello, 2001). Two dimensional point estimation requires distance measurements from three non-collinear points (Figure 3), whereas three dimensional point estimation requires distance measurements from four non-coplanar points required (Hightower & Borriello, 2001).

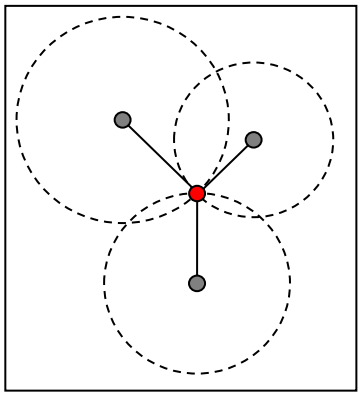

Figure 3. Lateration technique to estimate two dimensional location. Adapted from Hightower & Borriello (2001).

The lateration technique has three general approaches to measuring distance: 1) Direct, 2) Time of Flight (TOF), and 3) Attenuation.

1) *Direct measurement* relies on the physical movement. While straightforward, it is actually quite difficult to perform direct measurement of many objects because of the problem of cross-referencing movement between many-to-many relations and problems of isolating movement of an object in a complex environment.
2) *TOF measurement* uses the time that it takes for an object to travel a distance through a medium and calculates the distance by a known velocity of a signal such as ultrasound and light. TOF requires a clock with high resolution because of high velocity in signals (e.g., ultrasound: 344m/sec, light: 299,792,458m/sec). In addition, handling temporal agreement is another issue to consider. It is a challenging issue in TOF to discriminate signals arriving at an object by an indirect path caused by reflections in the environment with

obstructions such as buildings and trees because direct and reflected signals look identical. TOF has been widely applied in various LATs including GPS, the Active Bat Location System (Harter, Steggles, Ward, & Webster, 1999; Active Bat, 2009), and the Cricket Location Support System (Priyantha, Chakraborty, & Balakrishnan, 2000).

3) *Attenuation measurement* uses the intensity of a broadcast signal, which decreases as distance from the emission source increases. Given a distance-decay function correlating attenuation and distance for a type of broadcast and the original strength of the broadcast, it is possible to estimate the distance between the source and destination. The attenuation is, however, influenced by signal propagation issues such as reflection, refraction, scattering, and multipath, especially in indoor environments with many obstructions. This causes the attenuation to correlate poorly with distance, resulting in inaccurate and imprecise distance estimates, and generally the attenuation is less accurate than TOF. An example of attenuation-based LAT is the SpotOn ad hoc location system using low-cost tags (Higtower, Vakili, Borriello, & Want (2001); Hightower, Want, & Borriello (2000)).

Angulation: It uses angles to determine distance with direction. In general, two dimensional positioning requires two angles and one distance measurement (Figure 4), and a three dimensional position requires two angles, one distance, and one azimuth measurement. An example of angulation is the VOR (VHF Omni-directional Ranging) aircraft navigation system.

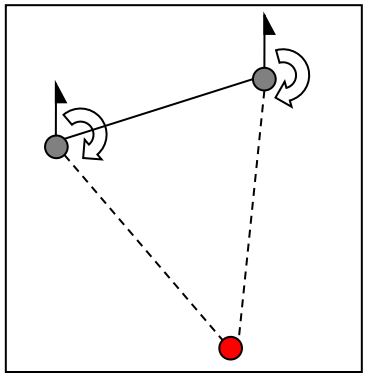

Figure 4. Angulation technique to estimate two dimensional location.

2.3.1.2 Proximity

A proximity location-aware technique detects an object when it is near (i.e., within limited range) to a known location (i.e., a physical phenomenon). Three general approaches are detecting physical contact, monitoring wireless access points, and observing automatic ID systems (Hightower & Borriello, 2001). Detecting physical contact is the most basic sort of proximity sensing, including pressure sensors, touch sensors, and capacitive field detectors (Hightower & Borriello, 2001). Monitoring when a mobile device is in range of one or more access points in a wireless cellular network is another implementation (Hightower & Borriello, 2001). Examples include the Active badge Location System (Want, Hopper, Falcao, & Gibbons, 1992) using infrared cells in an indoor environment. This can be automated in such a way that an object can be scanned and the identification information can be matched to a database to provide a location. Examples include credit card point-of-sale terminals, land line phones and computer login histories, and Radio Frequency Identification (RFID) badges

(Want & Russell, 2000). This may also involve use of mobile objects databases, which will (actively) update locations of objects in the database.

2.3.1.3 Scene analysis

Scene analysis can provide location awareness through pattern/feature recognition. This has the advantage of passively sensing movement. There are two types of scene analyses; static scene analysis and differential scene analysis. Static forms work by looking-up features in a dataset or database (data with some context) that maps them to object locations; differential forms work by studying the differences between scenes to estimate location (Hightower & Borriello, 2001).

2.3.2 Location Aware Systems

This section briefly introduces several location-aware systems. Table 1 represents a comparison of representative examples of LATs.

Table 1. Comparison among LATs. Adapted from Hightower & Borriello (2001).

Name	Properties			
	Technique	Attributes	Accuracy Precision	Scale
GPS	Radio, TOF, lateration	Physical, Absolute	1-10m, 95-99%	Minimum of 24 satellites cover worldwide
Active Badge	Diffuse infrared, Cellular proximity	Symbolic, Absolute	Room size	One base per room, Badge per base per 10 sec
Active Bats	Ultrasound, TOF, lateration	Physical, Absolute	9cm, 95%	One base per 10 sq m, 25 computations per room per sec
Cricket	Ultrasound, TOF and Proximity lateration	Symbolic, Absolute and Relative	4x4 ft regions, ≈100%	≈1 beacon per 16 sq ft
MSR RADAR	802.11 RF, scene analysis, triangulation	Physical, Absolute	3 - 4.3m, 50%	Three bases per floor

2.3.2.1 Outdoor environments

For applications in open, outdoor areas, satellite-based LAT is widely used. GPS is the classic example: it is an integrated system of satellites and ground radio receivers that allow for the triangulation of objects on the earth's surface relative to the ground and objects in space, making use of position, velocity, and time of delivery information. A radio signal is used to obtain the distance and position to each satellite, the GPS receiver computes its position using trilateration. Disregarding satellites beneath the earth, signals of which cannot be reachable, the satellites are always above the receivers so only three satellites would normally be required to estimate a 3D position (latitude, longitude, altitude); however, because the receiver is not synchronized with the satellite transmitters

and thus cannot precisely measure the time it took the signal to reach, a fourth satellite is required to have an agreement about time (Hightower & Borriello, 2001).

Standard GPS receivers can provide locations at accuracies of approximately 10 to 15 meters; however, it is important to notice that there are several possible sources of error inherent in these locations. Errors arise from signal degradation due to atmospheric effects, minor variations in the location of the satellites, inaccuracies in the timing clocks, errors in receivers, and variations in the reflection of signals (i.e., multipath effect) from local objects such as trees and buildings (Longley, Goodchild, Maguire, & Rhind, 2001).

The accuracy of measurement can be improved by using Differential GPS (DGPS). The DGPS signals were originally developed under the Selective Availability (SA) program, which degraded non-military use of GPS signals for security protection. Even though SA was permanently turned off in 2000, DGPS signals are still used today to enhance the accuracy of GPS units. The correction data is generated by a base reference station, which is a fixed GPS receiver located at an accurately known location. Errors are calculated by comparing the difference between the exact known location and the location calculated by satellite signals. The theoretical assumption is that receivers that are close together will show similar atmospheric errors. Potentially, DGPS can improve accuracy to allow locations to be determined to better than 1 meter; however, due to the assumption of a distance relationship, the accuracy of DGPS decreases with distance from the base reference station. The range to use DGPS is about 300 km

from the base reference station. Another system to improve the accuracy of measurements is Wide Area Augmentation System (WAAS). Similar to DGPS, WAAS uses a system of ground reference stations including two master stations positioned across the United States to provide necessary augmentations to the GPS Standard Positioning Service (SPS) navigation signal. Similar to GPS, there exist three other satellite based LATs: GLONASS (Global Navigation Satellite System) by Russian; GALILEO by European Union; and Beidou by China.

2.3.2.2 Indoor environments

For applications in the indoor environment, various indoor sensors have been developed in recent years such as *Active Badge* by infrared signal, *Active Bat* by ultrasound, *Cricket* by radio frequency and ultrasonic signals, *RADAR* by wireless LAN, and *ZPS* by ultrasounds.

Active Badge is the oldest indoor location sensor developed by Olivetti research laboratory (now at AT&T), Cambridge, UK (1989-1992) (Want, Hopper, Falcao, & Gibbons, 1992). It estimates location based on a cellular proximity system that uses diffuse infrared (IR) technology (Hightower & Borriello, 2001). *Active Bat* is an ultrasound-based location aware system developed by AT&T researchers. The system consists of a grid of ceiling-mounted receivers that receives ultrasound pulses emitted from multiple Bats (transmitters) attached to objects. It estimates the 3D physical location of Bats using the TOF lateration technique (Hightower & Borriello, 2001). In its experimental study, 720 ultrasounds receivers were placed throughout a building to cover an area of

around 1,000m^2 on three floors. The study showed that Active Bat system can determine the positions of up to 75 objects each second, accurate to around 3cm in three dimensions. Like the Active Bat system, Cricket uses ultrasound with radio synchronization and TOF and proximity lateration for symbolic location estimation (Priyantha, Chakraborty, & Balakrishnan, 2000). The system can accurately delineate 4x4 square-foot regions within a room. *RADAR*, developed by a Microsoft Research group, is a bulding-wide tracking system based on the IEEE 802.11 WaveLAN wireless networking technology (Bahl & Padmanabhan, 2000). The system uses signal strength and signal-to-noise ratio of signals that wireless devices send to compute the 2D position within a building for both lateration and scene analysis (Hightower & Borriello, 2001). The accuracy of RADAR is 4.3m for lateration and 3m for scene analysis respectively.

While above mentioned location aware systems is specifically targeting for indoor positioning, Local Positioning Systems (LBS) that use signals from cellular base stations and Wi-Fi access points have capability to both outdoor and indoor positioning. Positioning methods used in the former system include triangulation-based (e.g., E-OTD (Enhanced-Observed Time Difference), U-TDOA (Uplink-Time of Arrival)) and proximity (e.g., Cell ID) (Mishra, 2004). The accuracy ranges from 100m to several kilometers, which is the main drawback in the cellular-based positioning system. Wi-Fi based Positioning System (WPS) as the latter system uses radio signals from Wi-Fi access points and similar positioning estimation techniques used in cellular-based systems; however, because Wi-Fi access points are often deployed more densely in cities

than cellular towers, WPS is more accurate than cellular-based positioning systems. The accuracy ranges from 10 to 30 meters in urban areas in existing commercial systems such as Skyhook (Skyhook, 2008) and PlaceEngine (Rekimoto, Shionozaki, Sueyoshi, & Miyaki, 2006). In both cellular and Wi-Fi based systems, higher accuracy can be observed in dense urban areas where the density of cellular base stations and Wi-Fi access points is also high, whereas GPS works better in rural areas and less accurate in urban areas due to the multipath effect.

These LATs might be used to monitor not only just location but also to measure motion behavior by looking at a sequence of locations through time. Location itself is important information to study individual/collective human behavior in space and time such as location and its relation to space and time, spatial context and activity pattern, and location and its interaction with other humans. Motion-based analysis enables us to further examine human behavior such as transportation modes, motion behavior and its activity, and motion behavior and its interaction with physical environments as well as other humans.

2.4 Geographic Knowledge Discovery

Geographic Knowledge Discovery (GKD), a special case of Knowledge Discovery from Databases (KDD), is the human-centered process of extracting novel, interesting, and useful patterns from geo-reference data. Through the process of various data mining exercises, GKD is particularly useful for exploring spatio-temporal datasets collected by LATs, which are typically high-dimensional,

voluminous, and complex. GKD allows us to derive, for instance, meaning/context from movement/location data, while traditional spatial analysis is weak at handling such complex datasets.

Traditional spatial analysis methods often have limitations in handling voluminous datasets. Traditional analytics were developed when it was expensive to initiate large sampling exercises to collect data (usually manually), when the computing environment for processing these data was underpowered (in terms of computer processing abilities and the ability of databases to handle large volumes of data and large numbers of data queries); as a result, traditional techniques are not always ideal for analyzing conventional data, which are often massive in size as they are provided on an automated basis (Miller & Han, 2009). For example, traditional methods for measuring spatial dependency and heterogeneity effects (e.g., Moran's I and Geary's C for global analysis; Getis and Ord G, local version of I and G for local indicators of spatial analysis (LISA)) require approximately $O(n^2)$ in complexity. In addition, traditional statistical methods are confirmatory, meaning that they test data against a priori hypotheses; therefore, unlike exploratory research, they cannot discover unexpected or surprising information (Miller & Han, 2009).

There is an increasing demand for effective and efficient tools to extract hidden patterns, trends, and useful information and knowledge from spatio-temporal datasets by (automated) exploration; particularly for knowledge that is often buried in massive datasets that are also high-dimensional and complex (Mennis & Guo, 2009). In recent years, to address these challenges, there has

been a rise in interest in spatial/spatio-temporal data mining and GKD, specifically for theoretical investigation, algorithm and methodology development, and practice for the extraction of useful information and knowledge from massive and complex spatial databases (Andrienko & Andrienko, 1999; Guo, Peuquet, & Gahegan, 2003; Miller & Bridwell, 2009; Knorr & Ng, 1996).

2.4.1 Knowledge Discovery from Databases

Knowledge discovery through data-mining involves scouring datasets and databases, using some metadata, algorithm, or heuristic as a guide, usually benchmarking discovered patterns against a known ontology or template. KDD seeks interesting patterns that are hidden in very large databases. Such patterns are non-random properties and relationships are valid (a generalized pattern, not simply a data anomaly), novel (nontrivial and unexpected), useful (relevant), and understandable (interpretable) (Fayyad, Piatetsky-Shapiro, & Smyth, 1996). The KDD process usually involves multiple, connected steps, including data selection (e.g., selecting a subset of the records or variables), data preprocessing (data cleaning such as noise and outlier removal), incorporation of prior knowledge, data mining, visual representation, interpretation, and evaluation of the results (Fayyad, Piatetsky-Shapiro, & Smyth, 1996). Table 2 shows a possible classification of data mining-tasks and techniques.

Table 2. Data-mining tasks and techniques. Adapted from Miller and Han (2009).

Knowledge type	Description	Techniques
Classification	Predict the class label that a set of data belongs to based on some training datasets	Bayesian classification Decision tree induction Artificial neural networks Support vector machine
Clustering / Segmentation	Determining a finite set of implicit groups that describe the data	Cluster analysis
Association	Finding relationships among item-sets or association/correlation rules, or predict the value of some attribute based on the value of other attributes	Association rules Bayesian networks
Deviation	Finding data items that exhibit unusual deviations from expectations	Clustering and other data-mining method Outlier detection Evolution analysis
Trends and regression analysis	Lines and curves summarizing the database, often over time	Regression Sequential pattern extraction
Generalization	Compact description of the data	Summary rules Attribute-oriented induction

2.4.2 Geographic Knowledge Discovery

GKD, a special case of KDD, is the process of extracting hidden patterns, trends, and useful information and knowledge from massive and complex geo-referenced databases (Miller & Han, 2009). As with the data mining in KDD, spatial or spatio-temporal data mining in GKD encompasses various tasks and different techniques associated with the task. This section briefly introduces two representative tasks of spatial data mining: spatial classification and spatial clustering.

Classification deals with the assigning of things into categories. Spatial classification is a supervised classification technique that uses space as a container for data or that uses space to guide, calibrate, or validate the classification procedure. Spatial classification could also make use of a (spatial) training dataset to train the classification model, a validation dataset to validate the configuration, and a test dataset to evaluate the performance of the trained model (Mennis & Guo, 2009). Examples of classification methods include decision tree induction (Quinlan, 1986), naïve Bayesian classification (Domingos & Pazzani, 1997), artificial neural networks (Bishop, 1995), maximum likelihood estimation (Fisher, 1922), and support vector machine (Cortes & Vapnik, 1995). As extended from general classification methods, spatial classification is about finding rules to group spatial objects into predefined classes based on not only attribute values but also spatial attributes of the object (e.g., shape, extent) as well as spatial relationships to other objects. For example, Andrienko and Andriekno (1999) revealed spatial patterns of the classification rules based on decision tree algorithm, C4.5, using interactive map visualization.

While classification is a supervised learning approach, *clustering* is an unsupervised learning approach that partitions a selected set of data into meaningful groupings (clusters) so that items in the same group are similar to each other and different from those in other groups. Clustering can be based on combinations of non-spatial attributes, spatial attributes (e.g., shape, extent), and proximity of the objects or events in space, time, and space-time. Spatial clustering has been an active research field and many different clustering methods

have been developed. Major clustering methods can be generally classified into four categories; 1) partitioning method, 2) hierarchical method, 3) density-based method, and 4) grid-based method (Han, Lee, & Kamber, 2009).

1) Partitioning methods

Partitioning schemes are used to divide datasets into clusters using a set of formal guidelines. For example, a guideline might read as: "Given a database on *n* objects, a partitioning method constructs k($\leq n$) partitions of the data, where each partition represent a cluster. That is, it classifies the data into k groups with satisfying the following requirements: (1) each group must contain at least one object, and (2) each object must belong to exactly one group." (Han, Lee, & Kamber, 2009, pp. 154-155). Examples of partitioning methods are k-means (Lloyd, 1982), k-medoids (Kaufman & Rousseeuw, 1990), CLARANS (Ng & Han, 1994), and the EM algorithm (Dempster, Laird, & Rubin, 1977).

2) Hierarchical methods

Hierarchical schemes are used to classify data into hierarchical bins; i.e., each bin is related to the other in some tiled way. Examples include hierarchical decomposition that is agglomerative (bottom-up) or divisive (top-down) (Han, Lee, & Kamber, 2009, p. 155). An example of hierarchical method is BIRCH (Zhang, Ramakrishnan, & Livny, 1996).

3) Density-based methods

In many partitioning schemes, data are separated based on their differences in attributes; this can be considered as dividing data based on their attribute distance in some sort of attribute space. This can cause problems, because these techniques often have difficulties in detecting clusters of arbitrary shape (Han, Lee, & Kamber, 2009, p. 155). Density-based schemes are designed to overcome this problem: "The general idea is to continue growing a given cluster as long as the density (the number of objects or data points) in the neighborhood exceeds a threshold. Such a method is able to filter out noises (outliers) and discover clusters of arbitrary shape" (Han, Lee, & Kamber, 2009, p. 155). Examples of density-based methods are DBSCAN (Gaffney & Smyth, 1999) and OPTICS (Ankerst, Breuning, Kriegel, & Sander, 1999).

4) Grid-based methods

Grid-based schemes use grids as the template for partitioning data. This translates the data into a quantized space (the grid structure) and clustering is performed on that quantized space, with the advantage that processing time is often increased (Han, Lee, & Kamber, 2009, p. 156), (due to data compression, for example). Examples of grid-based methods are STING (Wang, Yang, & Muntz, 1997) and CLIQUE (Agrawal, Gehrke, Gunopulos, & Raghavan, 1998).

2.4.3 Trajectory Data Mining

Because of location-aware hardware in mobile objects (cars, people's pockets, devices, retail goods), there is increasing interest in performing data analysis over

trajectory datasets. This may be done by clustering, which is to group objects showing similar behavior and differentiate objects performing differently. For identifying trajectories of similar shapes, Gaffney & Smyth (1999) and Gaffney, Robertson, Smyth, Camargo, & Ghil (2006) have proposed a model-based clustering algorithm for trajectories. In these studies, a set of trajectories of hand movements in video streams (Gaffney & Smyth, 1999) and extratropical cyclones (Gaffney, Robertson, Smyth, Camargo, & Ghil, 2006) were clustered by introducing a probabilistic mixture regression model for such data and using the EM algorithm for clustering trajectories. In such an approach, each trajectory is considered as a whole; however, Lee, Han, & Whang (2007) argued that a trajectory may have a long and complicated path so that only some portions of trajectories show a common behavior, but the behavior is not common over the entire trajectory. As an alternative approach, Lee, Han, & Whang (2007) proposed a new clustering algorithm called TRACLUS which introduced a partition-and-group framework in order to discover clusters of sub-trajectories. In its framework, there are two phases; 1) the partitioning phase divides a trajectory into line segments by using the idea of minimum description length (MDL); and 2) a grouping phase clusters line segments that show similarity in some way, using a variation of DBSCAN (Lee, Han, & Whang, 2007).

Bogorny, Kuijpers, & Alvares (2009) pointed out that these approaches, however, suffer from four general problems that are essentially important for trajectory knowledge discovery:

1) "[they] focus on the mining step itself, basically considering the geometric properties of trajectory sample points, without taking into account the semantics of the data;

2) [they] do not cover the whole trajectory knowledge discovery process, which requires complex data preprocessing and post-processing tasks in order to generate meaningful patterns understandable by humans;

3) [they] do not consider the geography behind trajectories, which is the essential information to understand patterns in most application domains; and,

4) [they] do not provide preprocessing/transformation mechanisms to manipulate the data at different granularities (e.g. morning/afternoon, rush hours, weekday/weekend), which may be of fundamental importance in the knowledge discovery process." (p.1246)

In addition, Dodge, Weibel, & Forootan (2009) mentioned that, in fact, few trajectory data mining methodologies have been implemented and applied in practice. Weibel, Sack, Sester, & Bitterlich (2008) also argued that such trajectory data mining methodologies are currently being developed in a piecemeal/ad hoc fashion and have yet to migrate from research to demonstrate convincing social and commercial benefits. Furthermore, four research challenges in current-generation trajectory analysis schemes are identified (Cao, Mamoulis, & Cheung, 2009, pp. 405-406):

1) "a fundamental theory for modeling trajectory data and their access/analysis should be defined [...] (e.g., a set of typical analysis tasks should be defined and benchmark data should be provided for them);

2) it is necessary to develop a systematic framework that combines the dominant methods in managing and analyzing trajectory data;

3) some heuristics or models for setting and tuning parameters are required; and,

4) real applications impose additional requirements to data trajectory analysis (e.g., uncertainty in trajectory data due to translation delay or collection granularity)."

Thus, further exploration and investigation are required to advance the development of theory, methodology, and practice for the extraction of useful information and knowledge from massive and complex trajectory databases.

2.5 Complex Systems and Agent-Based Models

The complexity of an object's movement, or its dependencies on geographic context, may further complicate analysis. Complex systems, originally extended from the general system theory by von Bertalanffy (1968), can be understood through its important properties including openness, feedback, path-dependence, phase shift, non-linearity, emergence, and self-organization.

Complex systems are open and complex. 'Open' means that, in the systems, there are exchanges of matter, energy, and information with their

environment, while closed systems don't have interaction with environment. The exchanges, or interactions of input and output, have a feedback process whereby some portion of the output of a system is fed back to the input positively or negatively. Positive feedback increases and amplifies output exchanged between systems or system components and negative feedback has a reducing or dampening effect. A system's trajectory also has a property of path-dependence. The trajectories generated by such interactions are sensitive to their initial conditions or historical events; that is, qualitatively different/distinct trajectories emerge from the application of particular initial conditions. Such trajectories will be also locked-in to particular steady-state solutions like static and periodic dynamics. Dynamics of complex systems may also be non-linear, where a small perturbation may cause a large effect (colloquially called the butterfly effect), a proportional effect, or even no effect at all. Phase shifts are sharp transitions between different states of a system. Moreover, in contrast to a closed system where entities are in equilibrium status, complex systems, as an open system, may hold a non-equilibrium status or far from equilibrium status of their elements. In such systems, two important properties can be seen, emergence and self-organization. Holland (1998) notes that emergence centers on interactions that are more than a summing of independent activities, which involves nonlinear characteristics; and situations in which interactions described by simple rules can generate dynamical systems of surprising complexity. Such emergence is not analytically predictable from the attributes of internal components at lower levels. Self-organization is a process, in which the internal components of a system

increase in complexity without central controls. Such self-organizing systems typically display emergent properties. In addition to these properties in a complex system, a complex adaptive system has another important property, that is, adaptation. The system has a capacity to change and learn from experience.

Systems in reality are much like complex systems/complex adaptive systems (e.g., dynamics in climate, nervous systems, brain and immune system, stock markets, social insect and ant colonies, traffics and transportation networks, telecommunication infrastructures, and human migration and crowd dynamics). Stock markets, as a specific example, are comprised of millions of traders buying and selling in a bid to maximize their own individual profits. In such a system, individual investors act without any centralized control, yet their activities often lead to aggregate outcomes that are relatively efficient, as efficient as if they were controlled (notion of "invisible hand" by a Scottish economist Adam Smith, in 18^{th} century), that is, the system generates self-organization and adaptive behaviors.

The theoretical justification of adopting complex systems science in geographical research stems from the inherent spatiality of complexity. O'Sullivan, Manson, Messina, & Crawford (2006, p. 612) argued that "[b]ecause elements have some spatial configuration and interactions are not global but local, the spatial configuration of a system may be key to understanding and anticipating its behavior. The proper approach to space implied by this perspective involves close study of the local situational characteristics of physical locations, of interactions among neighboring locations, and of the flows along interactions

networks. Interactions among system elements are spatially structured in ways that contribute to the evolution of the spatial structure in which they play out." This argument follows a thesis introduced by Thrift: that complexity is "preternaturally spatial" (1999, p. 32). This inherent spatiality of complexity and the interplay between spatial configuration or pattern and process are similarly a central concern of the spatial sciences (O'Sullivan, Manson, Messina, & Crawford, 2006). Because "place is a complex web of social, economic, political and other relations, which are themselves spatially structured and configured over time" (O'Sullivan, 2004, p. 284), it is obvious to see a clear affinity between geography and complexity studies.

There have been myriad applications of complexity-based simulations to substantive questions in human geography as dynamic phenomena such as urban dynamics, residential mobility, retail behavior, traffic networks and crowd behavior (Benenson & Torrens, 2004; Batty, 2005). For example, urban development evolves over space and time as the result of micro-scale interactions of individual choices and actions (e.g., real estate transaction, residential mobility) taken by multiple agents such as households, businesses, developers, and governments (Alberti & Waddell, 2000). Such interactions affect urban and ecosystem structure, which will also fed back to the system, sometime lead to emergence of interesting phenomena, such as social segregation, urban growth and sprawl, and gentrification.

As a social segregation model, Schelling (1971; 1974) applied the idea of conflict and cooperation through game theory to social dynamics of segregation.

Schelling argued that residential segregation can be compatible with different micro-motives; and even mild segregationist preferences can bring about of residential segregation as a macroscopic phenomenon (Aydinonat, 2005). Thus, residential segregation could emerge as an unintended consequence of human action.

O'Sullivan (2002) developed micro-scale spatial modeling of gentrification using graph-based Cellular Automata. It was based on the demand-side theory, specifically using Smith's rent gap theory, which is the disparity between the potential ground rent level and the actual ground rent capitalized under the present land-use (Smith, 1979). Gentrification may be initiated when the gap is wide enough so that developers can cheaply purchase shells, physical housing structures, can pay the builders' costs and profit for rehabilitation, can pay interest on mortgage and construction loans, and can then sell the end product for a sale price that leaves a satisfactory return to the developer (Smith, 1979). The simulation outcomes successively generated such gentrification dynamics.

ABMs of pedestrian dynamics also described micro-behavioral complexity. For example, pedestrian crowd behaviors have been modeled using a social force model, which is based on physics and sociopsychological forces (Helbing & Molnár, 1997; Helbing, Molnár, Farkas, & Bolay, 2001). In these models, realistic collective crowd behaviors have been emerged from nonlinear interactions among individual pedestrians such as self-organization of lane formation and oscillatory flows through bottlenecks. These emergent behaviors

are not directly planned in simulations but the autonomous systems create such behaviors automatically.

Complex systems science can advance geographic researches because of natural affinity between properties in complex systems and real geographic phenomena. Particularly, the complex system's approach has a significant advantage over the traditional approach by looking at phenomena as detail, dynamic, and multi-scale behaviors with holistic approach, whereas the traditional approach views phenomena as relatively coarse, static, and inflexible at scale with reductionism approach. Another significant advancement can be a paradigm shift in studying geographic phenomena from prediction to experimentation and exploration by considering simulations as applied tools for evaluating plans and policies and supporting decision-makings. For example, simulation can be used as a tool for exploring and experimenting existing theories and ideas, and also simulation can be as an artificial laboratory for testing hypothesis with high degrees of realism and details (Brail & Klosterman, 2001).

To utilize these models, however, the model evaluation plays a critical role in complexity science. It involves three parts; 1) calibration, fine-tuning of the model to some dataset; 2) verification, testing the logic of model structure (e.g., see if models works in different software and show consistency); and 3) validation, measuring the goodness of fit between model and reality. Yet, most existing model evaluation approaches tend to be narrative and qualitative description thus more technical and quantitative approaches must be investigated.

In addition, most of available standard statistical methods are not directly oriented toward complexity. Manson (2007) discussed several challenges in terms of model evaluation of complexity. First of all, sensitivity analysis is useful for model use and evaluation. It also identifies tipping points and fine thresholds. Complex systems are sensitive in a sense that large and sudden shifts (phase shifts) in a system behavior can be a result from relatively small perturbations in inputs. Sensitivity is assessed by determining how incremental changes in input produce various outcomes and parameter sweeping is a typical method for evaluating a simple model (Manson, 2007); however, in complex systems, because small changes may produce large difference, sensitivity analysis may be a difficult issue. In addition, the characteristic of non-linearity also makes sensitivity analysis difficult because output behavior is not proportional to at least some potion of inputs and more it includes interactions and feedback effect. Therefore, it requires sophisticated test design to identify tipping points and fine thresholds.

Second, in complex systems, macro-scale outcomes of emergence are results from micro-scale interactions among internal components, and the emergence is not analytically tractable from the attributes of those internal components; therefore, it is difficult to explain causal relationships of emergence among multi-scale elements in a system (Manson, 2007). It is important to know that very different combinations of micro-state behaviors can produce seemingly identical macro-state behaviors (Sawyer, 2002).

Third, the model evaluation should not only focus on patterns but also processes of complexity. Research of geographic complexity can easily evaluate a system considered complex if it merely exhibits certain patterns of complexity, whereas the conflation of pattern and process is one of the most exciting aspects of complexity research because hallmark patterns of complexity may lend insight into complex processes (Manson, 2007).

Fourth, the inductively model calibration may be problematic. For example, there are many land use models of CA and ABM that link theory to models and link the models to reality by calibrating them against empirical observations through full parameter enumeration. The shortcoming of model calibration in this manner is that the model may not apply to situations beyond those found during the inductive calibration stage (Hodges & Dewar, 1992).

Finally, absolute validation and verification of models of natural systems is impossible because the models are simplifications of open systems, whereas closed systems can be fully validated (Manson, 2007). The same argument extends to human-environment and social systems because they are obviously 'open' systems (Batty & Torrens, 2005). Therefore, models can only be evaluated subject to four kinds of uncertainty: theoretical, empirical, parametric, and temporal.

In summary, while it would be useful to represent the complexity of geographic systems in which movement manifests, doing so with existing methods is difficult and new techniques are needed.

Chapter 3

RESEARCH OBJECTIVES

A review of the literature highlights three key research challenges for spatio-temporal analysis and modeling of human movements; 1) a lack of space-time analysis tool; 2) a lack of empirical data analysis for spatio-temporal context awareness of human movements; 3) a lack of studies about evaluating simulation model of human movements; and 4) challenges in handling complexity, particularly emergence across scales. This dissertation research aims to investigate all four research challenges by conducting three studies on space-time analysis and modeling. Research objectives for three studies are as follows.

Study 1

- Developing an integrated spatio-temporal data exploration tool to represent spatio-temporal pattern and process of mobile objects.
- Incorporating the framework of time geography for qualitative visualization of mobile objects.
- Incorporating quantitative representation of mobile objects.

Study 2

- Developing a trajectory data mining methodology for context awareness of human movement.
- Generating theoretical movement data by random walk models.
- Collecting data of human spatio-temporal movements by GPS.

- Analyzing movement dataset with spatio-temporal data exploration tool and trajectory data mining method.

Study 3

- Developing an agent-based simulation model of pedestrian evacuation dynamics to explore pedestrian complex behaviors.
- Quantitatively and qualitatively extracting pedestrian complex behaviors using spatio-temporal data exploration tool and trajectory data mining method for evaluation of simulation models.

A major research task in GIScience is to provide methods to analyze and understand the spatio-temporal patterns, processes, and behaviors of mobile objects, as extraction of useful information and knowledge about dynamic and mobile phenomena. A key challenge is to analyze and visualize a large dataset of multiple mobile objects for better understanding of movement behaviors, their interactions, and collective behaviors through space and time. The first study develops an integrated spatio-temporal data exploration tool to represent spatio-temporal patterns and process of mobile objects and seeks to contribute to the challenge. The tool uses time geography to integrate both quantitative and qualitative representations of mobile objects. It incorporates the quantitative representations of motion behavior including basic motion descriptors (e.g., velocity, acceleration, orientation, length, and sinuosity), fractal dimension, directional distribution, and Lévy metrics, and the 3D visualization of space-time

trajectory as a qualitative approach. These provide an interactive environment for human activity exploration and help to visualize, quantify, and analyze geographical patterns and tendencies in relation to time.

The second study offers insight into the research challenge of space-time data and context aware trajectory analysis. Applying the tool developed in the first study and extending it by adding a trajectory data mining method, it explores spatio-temporal pattern and process of movements. With the tool, the second study specifically aims to tackle three research challenges; 1) how to characterize and generalize massive trajectories to extract interesting patterns; 2) how to explain behavioral contexts of trajectories by those extracted patterns; and 3) how to visualize extracted patterns to overview and compare patterns and trends in space and time. To examine the capability of the toolkit for extracting interesting patterns, explaining behavioral context, and visualizing extracted patterns, two datasets of mobile objects were analyzed. The first is theoretical movements generated by three random walk models. The behaviors of these models are known because they are explicitly defined by mathematical expressions; therefore, it is useful to examine how the proposed toolkit answers three research challenges. The dataset consists of mixed trajectories simulated by three random walk models; Brownian Motion (BM), Correlated Random Walk (CRW), and Lévy flight. As the second dataset, GPS tracks of real movement were used to test the data mining scheme on real-world data.

The third study contributes to the research challenge of evaluating an Agent Based Model (ABM) of human movement. A key research challenge is

model validation, which is a difficult task when systems in the real-world as well as generated by ABM exhibit complex behaviors, such as feedback, path-dependence, phase shift, non-linearity, emergence, adaptation, and self-organization. A specific challenge in model validation is which aspects of the model behavior are to be compared with empirical data. Complex behaviors cannot be simply examined by looking at global statistics, but it is necessary to consider spatio-temporal process and behaviors across various scales. This study proposes a new analytical framework for evaluating ABMs by applying the developed tool in the first and second studies. It utilizes a trajectory data mining technique that uses trajectories of mobile objects from real-world and ABMs as input datasets, partitions the trajectories into sub-trajectories, and identifies behavioral clusters based on their motion characteristics. The extracted patterns will be compared and visualized within the concept of time geography to exploratory investigate spatio-temporal patterns and trends. To examine the proposed framework, this study develops an ABM of pedestrian crowd dynamics under evacuation in a four-way intersection using the social force model. Then, crowd dynamics under four scenarios are compared in order to examine model behaviors as well as to investigate the effect of different designs of intersection to evacuation dynamics.

Chapter 4

SPATIO-TEMPORAL ANALYSIS AND VISUALIZATION OF MOBILE OBJECTS

4.1 Overview

Classic The first study is concerned with collecting data regarding mobile objects, and "making sense" geographically of those data, using spatio-temporal analysis and visualization. Later, these data will be used in models of human movement. A major research task in GIScience is to provide methods to analyze and understand the spatio-temporal patterns, processes, and behaviors of mobile objects, as extraction of useful information and knowledge about dynamic and mobile phenomena. In particular, a key challenge is to analyze and visualize a large dataset of multiple mobile objects for better understanding of movement behaviors, their interactions, and collective behaviors through space and time.

The specific contribution of this study is to introduce an integrated spatio-temporal data exploration toolkit to represent spatio-temporal patterns and process of multiple mobile objects. The toolkit integrates both quantitative and qualitative representations of mobile objects utilizing the framework of time geography. The quantitative representation includes quantifications of mobile objects by basic motion descriptors (e.g., velocity, acceleration, orientation, length, and sinuosity), fractal dimension, directional distribution, and Lévy metrics, whereas the qualitative representation incorporates the 3D visualization of space-time trajectories. A case study demonstrates the functionality of the toolkit by

analyzing pedestrian crowd dynamics under evacuation scenarios generated by an ABM.

4.2 Related Works

One approach in GIScience to investigate mobility data is to employ the concepts of Hägerstrand's time geography and its central principles/methods of space-time paths (STPs) and space-time prisms (Hägerstrand, 1970). In time geography, individual movements over time space-time trajectories reside in a 3D space where the X and Y axis represent geographic positions and the Z axis, a perpendicular dimension, represents time. Space-time trajectories provide an event-oriented framework for analyzing individual's activities based on spatial and temporal change with space and time constraints. A space-time prism describes the extent in space and time that an individual can access under a specific set of constraints. The 3D visualization of space-time trajectory and prism in GIS provides an interactive environment for human activity exploration and helps to visualize, quantify, and analyze the geographical patterns and tendencies in relation to time.

Considerable efforts have been made to develop analytical methods for time geography, including the formalization of conceptual frameworks and visualization techniques (Miller, 1991; Hornsby & Egenhofer, 2002; Kraak & Koussoulakou, 2004; Yu & Shaw, 2007). Recently, several analytical tools employing time geography concepts have been implemented in GIS environments (Kwan, 2000a; Kapler & Wright, 2004; Yu & Shaw, 2008; Miller & Bridwell,

2009); however, limitations have been acknowledged. First, quantifications of space-time trajectories have not been incorporated into visualization of the time geography framework effectively. Second, visual inspection of collective mobility patterns reaches its limits if numbers of mobile objects and lengths of space-time trajectories increases (Kwan, 2000a; Shaw, Yu, & Bombom, 2008). These limitations are partially due to the weakness of conventional GIS software to handle volumetric 3D objects.

Various quantifications can be computed to describe the behavior of mobile objects such as speed, acceleration, turning angle, displacement (i.e., the beeline distance between two points), travel path (i.e., the total length of a trajectory), and straightness index (i.e., the ratio of the traveled path and displacement) (Benhamou, 2004; Laube, Dennis, Forer, & Walker, 2007; Dodge, Weibel, & Lautenschütz, 2008). Quantification is an important precondition to compare either the motion of individuals or to make comparisons between different kinds of mobile objects. In addition to basic motion descriptors that describe the properties of movement, quantitative analyses such as fractal analysis and distance/directional distribution analysis that contextualize movement are useful for exploring a general understanding of the basic laws governing the object's motion.

The idea of the fractal has been applied for measuring tortuosity of movement paths. Mandelbrot (1967), who coined the term *fractal*, spread the idea of fractal geometry. In standard Euclidean planes, 0-dimension refers to point, 1-dimension refers to length, 2-dimension refers to area, and 3-dimension refers to

volume. In contrast, the fractal dimension is non-integer and always greater than the ordinary Euclidean dimension for a given object. The fractal dimension provides a measure of how densely an object fills space or how many parts of an object are observed as measurement resolution becomes finer. In the case of a linear feature including movement paths, the fractal dimension lies between 1 and 2, where 1.0 represents a straight path and 2.0 indicates that a path is so tortuous as to completely fill a plane. Fractal analysis has been used in various types of studies of animal movements and habitats, for example, the landscape perceptions of grasshoppers (With, 1994), habitat selection at different spatial scales of marten (Nams & Bourgeois, 2004), and scale-dependent movements of seabirds (Fritz, Said, & Weimerskirch, 2003). Because of its attention to geometries between dimensions, the metric is particularly appropriate for examining various features of movement paths in relation to various spatial scales.

To explore the statistical properties of objects' mobility patterns, the statistical distribution of displacement has often been examined. For example, exploring whether the displacements of mobile objects follow a normal distribution or a power-law distribution could support general understanding of the basic law or process governing an object's motion. In particular, biologists have studied whether the distribution of animal movement exhibits a Lévy flight pattern (Fritz, Said, & Weimerskirch, 2003). A mathematical concept of a Lévy flight is a special case of random walk, in which the distribution of distances in each step has long-tail probability. The distribution used is a power law in which the probability of large steps of size *D* might fall off in proportion to d^{γ}, with γ

being a number somewhere between 1 and 3. A Lévy flight is considered as an efficient strategy for foraging behaviors in biology. Recent studies showed evidence that some animals exhibit Lévy flight patterns; for example, monkeys (Ramos-Fernandez, Mateos, Miramontes, Cocho, Larralde, & Ayala-Orozco, 2003), sharks, turtles, and penguins (Sims, et al., 2008). Moreover, Brockmann, Hufnagel, & Geisel (2006) tracked dollar notes moving through the United State, and found that the distribution of distances travelled over a short time follows a power law with a γ equal to about 1.6. The result indicates that human travel patterns follow Lévy flight because money is carried by individuals so that its dispersal is a proxy for human movement (Brockmann, Hufnagel, & Geisel, 2006). Furthermore, movement data collected by anonymized mobile-phone for more than 100,000 people over a 6-month period follows the Lévy flight pattern (González, Hidalgo, & Barabási, 2008).

Directional statistics (Batschelet, 1981; Mardia & Jupp, 2000) allow for the exploration of directional patterns of mobile objects. Directional autocorrelation of movement is a key issue in investigating turning angle distributions. For example, behavioral ecologists may examine constancy patterns by investigating directional persistence in turning angle distributions (Turchin, 1998). A study by Schmitt and Seuront (2001) showed that some copepod species show intermittently constant straight sequences in their foraging behavior. As another example, desert ants, after having performed a circuitous foraging journey, find reliably the most direct way straight back to their nest from a distance of up to 100 m (Knaden & Wehner, 2003; 2004).

To visualize motion descriptors of space-time trajectory requires representing 4-dimensional information (i.e., x and y for space, z for time, and a scalar value for a motion descriptor). In addition, because visual inspection of the collective mobility patterns are limited by the number and lengths of space-time trajectories, advanced visualization techniques are needed to better capture collective movement behaviors in space and time. Most traditional GIS, however, handles geographic data in 2D or 2.5D (i.e., single value of Z coordinate), but have difficulty in handling 3D data (i.e., multiple Z coordinates) and beyond (Abdul-Rahman & Pilouk, 2008).

4.3 Methodology

To facilitate the application of movement analysis to large data-sets (whether collected from LATs or generated in simulation), this study builds a Space-Time Analysis toolkit. The toolkit will be developed for building a STP, which is an individual trajectory between two space-time anchors, using a two-dimensional plain to show geographical positions and use perpendicular dimension to represent time. These representations are accessible via a spatial database so that large data can be organized and queried using the ideas of time geography. Providing the 3D visualization of STPs helps to qualitatively and quantitatively analyze the spatio-temporal patterns and tendencies for movement data. In terms of quantitative representation, each individual trajectory can be described and characterized as measurable motion descriptors including velocity, acceleration, direction, length, and sinuosity. Such quantification is an important precondition

to compare either the motion of individuals or between different kinds of mobile objects. In addition, the toolkit incorporates fractal dimension analysis, distance/directional distribution analysis, and Lévy metrics, which are useful for exploring general understanding of the basic law governing the object's motion. These multi-dimensional quantifications of trajectory are also visualized as STPs using color representation and enhanced by stream tubes representation. Furthermore, the toolkit employs Space-Time Kernel Density Estimation (STKDE) and volume rendering techniques to better capture collective movement behaviors.

4.3.1 Quantitative Analysis of Mobile Objects

The Space-Time toolkit incorporates a set of spatial and space-time analysis methods for contextualizing movement, measuring movement, and comparing movement. These will be developed around 1) velocity and acceleration, 2) sinuosity, 3) fractal dimension, 4) power-laws, and 5) directional statistics. In each case, a mixture of visual and empirical metrics/schemes is developed.

4.3.1.1 Velocity and acceleration

Velocity and acceleration show general properties of movement relative to a fixed point or to a prior speed. These properties can differentiate motion behaviors; for example, velocity can explain modes of mobile objects such as walk, run, drive, and stop/stay, whereas the change in acceleration can describe phase shifts of such motion behaviors in relation to speed.

For an object's two dimensional vector moving from point P to point Q, the displacement of the mobile object is the change in the position vector *r*, given the *x* and *y* component of Δr as Δx and Δy, and Δt referring to the duration of the described motion (Figure 5).

$$\Delta x = x_Q - x_P$$

$$\Delta y = y_Q - y_P$$

$$\Delta t = t_Q - t_P$$

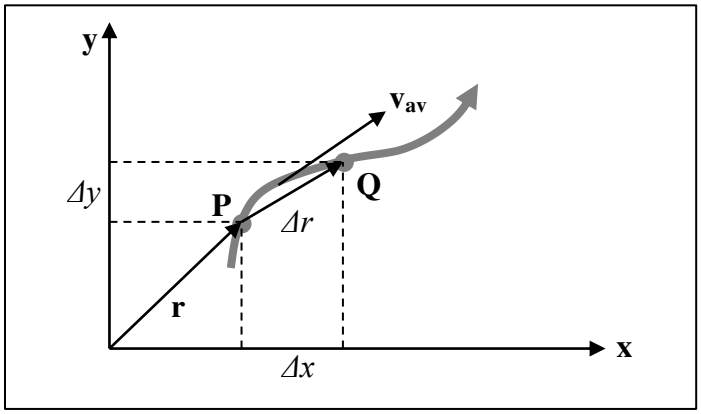

Figure 5. Velocity of a mobile object. Adapted from Sears, Zemansky, & Young (1987).

In kinematics, the average velocity v_{av} is defined to be the vector quantity equal to the displacement divided by the time interval (Sears, Zemansky, & Young, 1987) as follows.

$$v_{av} = \frac{\Delta r}{\Delta t}$$

The average velocity is a vector quantity having the same direction as Δr and a magnitude equal to the magnitude of Δr divided by Δt. The magnitude of Δr is the straight line distance from P to Q.

The instantaneous velocity, the velocity at a specific point in the trajectory at some instant time, is defined in magnitude and average velocity when P is taken closer and closer point Q.

$$v = \lim_{\Delta t \to 0} \frac{\Delta r}{\Delta t} = \frac{dr}{dt}$$

The average acceleration, a_{av} of an mobile object from Point P to Point Q is defined as the vector change in velocity, Δv, divided by elapsed time Δv (Sears, Zemansky, & Young, 1987).

$$a_{av} = \frac{\Delta v}{\Delta t}$$

The instantaneous acceleration, a, of an mobile object refers in analogy to instantaneous velocity to its acceleration at some point of its trajectory at some instant of time. It is defined in magnitude and direction as the limit approached by the average acceleration when point Q approaches point R and Δv and Δt both approach 0.

$$a = \lim_{\Delta t \to 0} \frac{\Delta v}{\Delta t} = \frac{dv}{dt}$$

4.3.1.2 Sinuosity

The measurement of sinuosity describes tortuosity, a property of a movement path being tortuous or crooked. The sinuosity of trajectories has been studied in last

two decades largely in the field of biology and ecology to investigate the animal's movement path in relation to its habitats. Batschelet (1981) promoted the use of a simple and intuitively appealing straightness index, which is the ratio of the straight distance between the start and end points of the path (*D*) and the distance measured along the path (*L*). Relative sinuosity can show how exaggerated a path is compared to another path, which might be a result of environmental complexity, for example.

$$\tau = \frac{D}{L}$$

The range of the straightness index is between 0 and 1, where 1.0 represents a straight path. The straightness index has been applied to study the migration mechanism of sea turtles (Pari, Luschi, Akesson, Capogrossi, & Hays, 2000) and the flight pattern and foraging behavior of free-ranging wandering albatrosses (Weimerskirch, Bonadonna, Bailleul, Mabille, Dell'Omo, & Lipp, 2002). According to Benhamou (2004), however, there is no theoretical study yet attempted to determine the reliability of the index as a measure of the orientation efficiency.

4.3.1.3 Fractal dimension

Fractal dimension can show 1) how much a path fills space and can therefore provide another measure of relative sinuosity, and 2) the likelihood of a movement path to retain its shape over scale. As the straightness index looks at sinuosity of a movement path at a global scale, the fractal dimension metric can

examine relative sinuosity at different spatial scales; for example, it can quantitatively measure a movement path which may be composed of goal-oriented movement at macro-scale (e.g., work to home) and wandering movement at micro-scale (e.g., shopping on the way to home, wandering of pedestrian on the street due to high crowd density).

To estimate the value of fractal dimension, *D*, a conventional approach is the dividers method, which is based on the empirical studies of coastlines and was used by Mandelbrot (1967) to quantify curves whose fractal dimensions were greater than one. The basis of the method is to measure the length of the curve by approximating it with a number of straight-line segments, called steps (Boschetti, Dentith, & List, 1996). The calculated length of the curve is the product of the number of steps and the length of the step itself. As the step size is decreased, the straight-line segments can follow the curve more closely, smaller-scale structure becomes more significant, and the calculated length of the curve increases. The mathematical form is expressed as follows.

$$L(\lambda) \propto \lambda^{(1-D)}$$

where: λ is the step length, $L(\lambda)$ is the length of the curve based on the unit measurement length λ, and *D* is the fractal dimension of the curve. Plotting the logarithm of the step length versus the logarithm of the corresponding curve length, a Mandelbrot-Richardson plot is obtained. The slope of a line fitted to these points is related to the degree of complexity of the curve being analyzed. This slope is related to the fractal dimension by the equation,

$$D = 1 - S$$

where *D* is the fractal dimension and *S* is the slope of the line (Kennedy & Lin, 1986). The slope of the Mandelbrot-Richardson plot is equal to, or less than, 0 so that, in the case of a curve, the fractal dimension is between 1 and 2. This yields one overall estimate for *D* over a range of scales.

4.3.1.4 Power-law distribution

The power-law/long-tail distribution of displacement can be used to examine the statistical property of objects' mobility patterns. In particular, identifying the power-law relationship can describe general motion behavior such as Lévy flight pattern.

To identify power-law behavior, we can examine if a histogram of a quantity appears as a straight line when plotted on logarithmic scales. There are three plotting methods; a normal histogram, 2) a histogram with logarithmic binning, and 3) a plot with a cumulative distribution function (Newman, 2005). While the first two approaches have noise in the tail distribution, a plot using a cumulative distribution function is a superior method. To estimate the exponent of power-law distribution γ, one way is to fit the slope of the line in plots, which is the most commonly used method (e.g., a least-squares fit of a straight line). However, it is known to introduce systematic biases into the value of the exponent (Goldstein, Morris, & Yen, 2004). An alternative method is to employ maximum likelihood methods as follows (Newman, 2005).

$$\gamma = 1 + \mathrm{n}\left[\sum_{i=1}^{n} \ln \frac{x_i}{x_{min}}\right]^{-1}$$

An estimate of the expected statistical error σ on the estimation by maximum likelyhood method is given by

$$\sigma = \sqrt{n}\left[\sum_{i=1}^{n} \ln \frac{x_i}{x_{min}}\right]^{-1} = \frac{\gamma - 1}{\sqrt{n}}$$

4.3.1.5 Directional statistics

Directional statistics use descriptive (and usually visual) methods for illustrating the general directional tendency in data. For example, to explore turning angle distribution, radar plots visualize the turning angle distributions around the compass card in a very illustrative way (Figure 6).

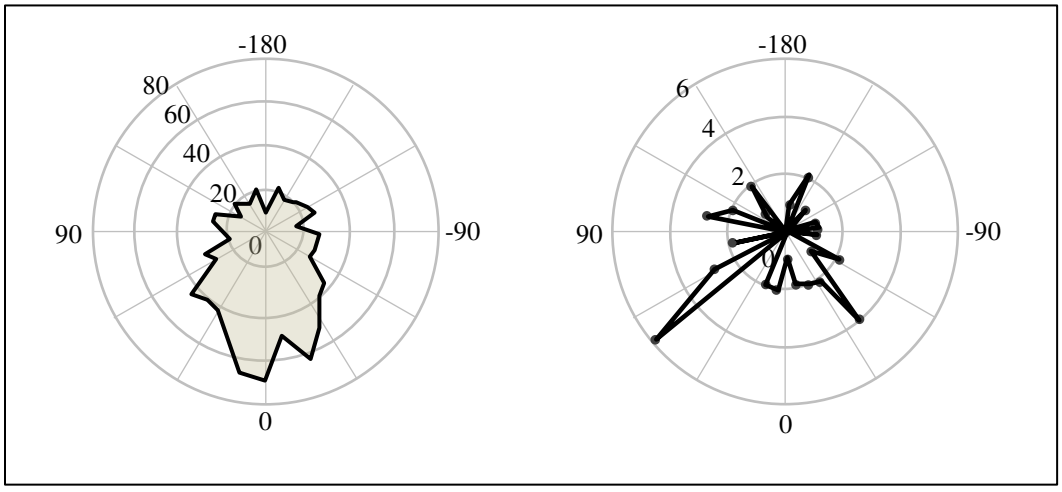

Figure 6. Radar plots. Direction change frequency distribution of Porcupine Caribou Herd (PCH) sample (Left) and direction change frequency distribution of caribou individual Blixen (right). Adapted from Laube & Purves (2006).

In addition, the directional pattern on movements can be examined by calculating the directional mean and circular variance. The directional mean ($\bar{\theta}$) is calculated as follows:

$$\bar{S} = \frac{1}{n}\left(\sum_{i=1}^{n} \sin\theta_i\right)$$

$$\bar{C} = \frac{1}{n}\left(\sum_{i=1}^{n} \cos\theta_i\right)$$

where $\bar{S}$ is the mean sine, $\bar{C}$ is the mean cosine, n is the total number of vectors.

$$\bar{\theta} = \tan^{-1}\left(\frac{\bar{S}}{\bar{C}}\right), \qquad if\ \bar{S} > 0\ and\ \bar{C} > 0;$$

$$\bar{\theta} = \tan^{-1}\left(\frac{\bar{S}}{\bar{C}}\right) + \pi, \qquad if\ \bar{S} > 0\ and\ \bar{C} < 0;$$

$$\bar{\theta} = \tan^{-1}\left(\frac{\bar{S}}{\bar{C}}\right) + 2\pi, \qquad if\ \bar{S} < 0\ and\ \bar{C} > 0;$$

The circular variance (S_v) is calculated as follows:

$$\bar{R} = \sqrt{(\bar{S}^2 + \bar{C}^2)}$$

$$S_v = 1 - \left(\frac{\bar{R}}{n}\right)$$

where $\bar{R}$ is the mean resultant length.

The range of circular variance is from 0 to 1; when the value is close to 0 this indicates that all vectors go generally in the same directions and, when it is close to 1 it indicates that vectors go in various directions. These two descriptive

statistics, the directional mean and the circular variance, exhibit the central tendency and the variability of the directions for an individual's movement respectively.

Furthermore, the directional autocorrelation in trajectory can be compared to random walk models, where a positive autocorrelation, or directional persistence, describes that the direction of the current move affects the direction of the next move. For a positive directional autocorrelation, the turning angles are concentrated around zero (Turchin, 1998). Bergman, Schaefer, & Luttich (2000) investigated the directional autocorrelation of two differently behaving caribou herds, and could distinguish a migratory and a stationary herd type.

4.3.2 Qualitative Visualization of Mobile Objects based on Time Geography

The qualitative approach that I introduce employs the framework of time geography to visualize movement of mobile objects enhanced by quantifications of space-time trajectories; thus, it is able to capture behaviors of mobile objects not only spatially but also temporarily. The toolkit has two methods to represent mobile objects, STP and STKDE.

4.3.2.1 Visualization of space time path

The first method is a visualization of an individual trajectory as a STP, which is composed of a sequence of vertices represented in a 3D space-time aquarium.

$$STP_{(i)} = P_{(1)}, P_{(2)}, P_{(3)}, \dots, P_{(j)}$$

where, *i* is an individual mobile object, *j* is a number of vertices, and $P_{(j)}$ is a 3D space time point (x, y, time). A STP is also composed of a sequence of segments. Figure 7 illustrates a single STP and multiple STPs (n=10) respectively.

A STP can be also described as a sequence of segments.

$$STP_{(i)} = S_{(1)}, S_{(2)}, S_{(3)}, \dots, S_{(k)}$$

where, k is a number of segments ($k=j-1$), and $S(k)$ is composed of two vertices, $P(k)$ and $P(k+1)$. Each segment (S(k)) possesses a set of scalar values of basic motion descriptors, M(k), calculated by quantitative analysis (length, duration, average velocity, average acceleration, and direction). STP visualization can be enhanced by the use of color based on these scalar values (Figure 8). The scalar value in Figure 8 is an average velocity of each segment. Furthermore, a stream tube representation can emphasize STP visualization. A stream tubed STP is wrapped with a tube whose radius is proportional to a scalar value. In Figure 9, the radius of a tube is proportional to the inverse of average velocity magnitude of a segment so that a fat tube represents slow and a thin tube represents fast respectively. In Figure 10, the radius of a tube is proportional to the average acceleration magnitude of a segment so that a fat tube represents high acceleration and a thin tube represents low acceleration. It provides better understanding of changes in motion descriptors in space and time; however, it will be difficult to understand if there are many STPs due to the occlusive effect.

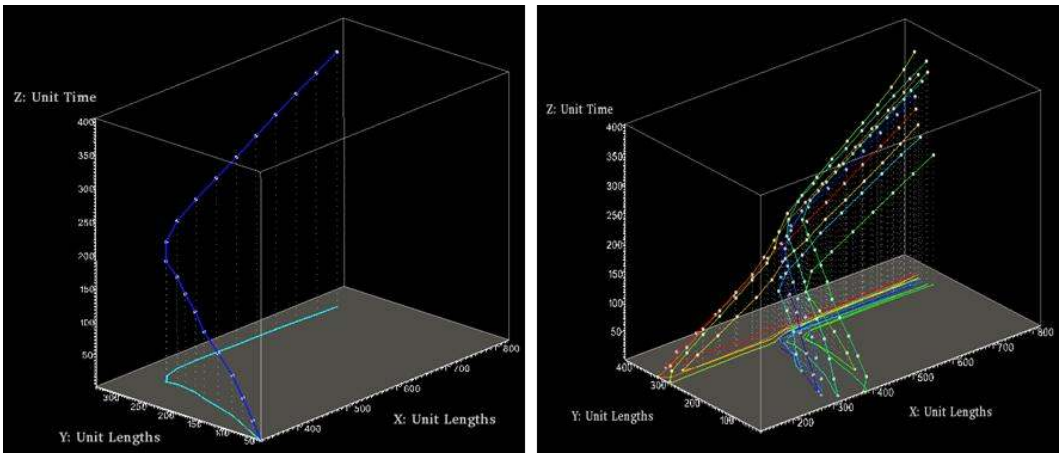

Figure 7. STP (left: single path, right: multiple paths colored by path ID).

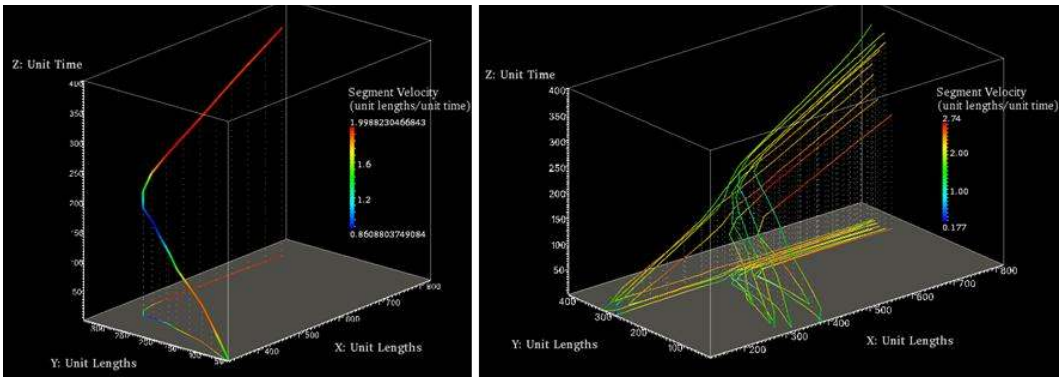

Figure 8. STPs colored by velocity value (left: single path, right: multiple paths).

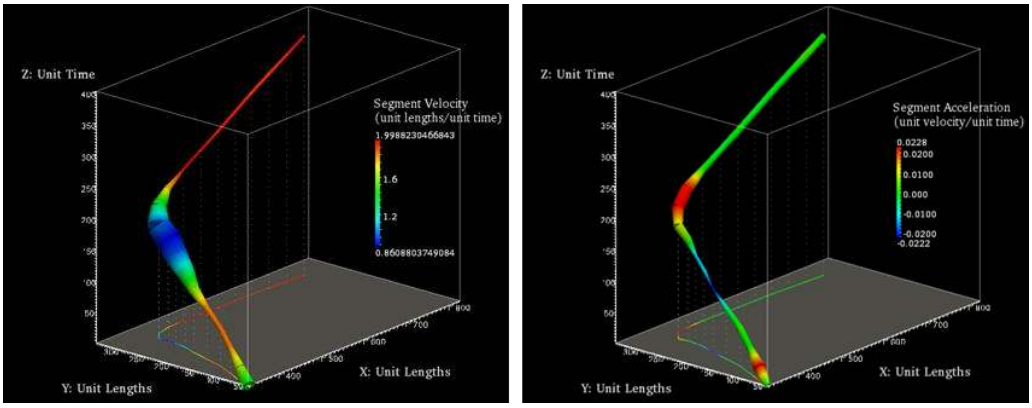

Figure 9. Stream tubed STP (left: single path, Inverse velocity, right: acceleration).

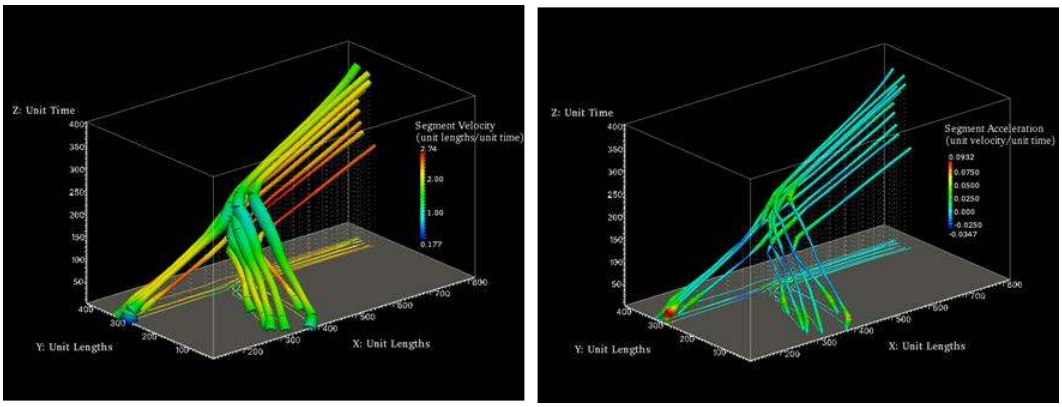

Figure 10. Stream tubed STPs (left: multiple paths, Inverse velocity, right: acceleration).

4.3.2.2 Space-Time Kernel Density Estimation for mobile objects

The second approach employs STKDE and a volume rendering technique to geovisualize the density of mobile objects in a 3D space-time aquarium. While the STP approach can create difficulty in understanding movement behaviors with multiple mobile objects due to the occlusion effect, the STKDE approach with the volume rendering technique is a useful for identifying space-time hot/cold spots of large and complex movement behavior such as in a crowd.

In the two-dimensional kernel density method, an estimate of probability density at the point (x, y) is given by;

$$\hat{f}(x,y) = \frac{1}{nh_1^2}\sum_{i=1}^{n} K_1\left(\frac{x-x_i}{h_1}, \frac{y-y_i}{h_1}\right)$$

where K_1 is a kernel function defined over 2-dimensional space, h_1 is the bandwidth of the estimate (i.e., a search space radius around the point (x,y) that

controls the smoothness of the estimate), and n is the number of observations of the form (xi, yi) for *i*=1…n.

The two-dimensional kernel density estimation can be extended to investigate spatio-temporal datasets in a 3D space. STKDE attempts to estimate the probability density function on point events distributed in 3-dimensional space where x and y values represent longitude and latitude and z value express time. Like a kernel density estimate in 2D space, a 3D kernel density estimation of probability density at point (x, y, t) is given by;

$$\hat{f}(x,y,t) = \frac{1}{nh_1^2h_2}\sum_{i=1}^{n} K_1\left(\frac{x-x_i}{h_1}, \frac{y-y_i}{h_1}\right) K_2\left(\frac{t-t_i}{h_2}\right)$$

where K_2 is a kernel function defined over time with bandwidth h_2 and n is the number of observations of the form (x_i, y_i, t_i) for i=1…n. In kernel density estimation methods, selection of kernel function as well as bandwidth influences the quality of a density estimate. However, Scott (1992) mentioned that the quality of density estimation is primarily determined by the choice of bandwidth, and choice of kernel is not crucial.

Kernel functions that are commonly used are, for example, uniform, triangular, biweight, gaussian, and Epanechnikov (Silverman, 1986). The toolkit developed in this study implemented Epanechnikov kernel function (Epanechnikov, 1969) as follows.

$$K_1(u,v) = \frac{2}{\pi}\{1-(u^2+v^2)\}, \qquad if\ (u^2+v^2) < 1, \qquad otherwise\ 0.$$

$$K_2(w) = \frac{3}{4}(1-w^2), \qquad if\ w^2 < 1, \qquad otherwise\ 0.$$

The density is in units of number of points per square length of unit per time of unit. In addition, the point density value can be magnified by a scaling factor (α), which can be a scalar value of motion descriptors. In this case, the above kernel functions are adjusted as follows.

$$K_1'(u,v) = \frac{2}{\pi}\{1-(u^2+v^2)\}\alpha, \qquad if\ (u^2+v^2) < 1, \qquad otherwise\ 0.$$

$$K_2'(w) = \frac{3}{4}(1-w^2)\alpha, \qquad if\ w^2 < 1, \qquad otherwise\ 0.$$

The density is then in units of the scaling factor used per square length of unit per time of units.

To obtain optimal bandwidth, much discussion has seen and many techniques have been proposed (Silverman, 1986; Scott, 1992). In the field of GIScience, as rules of thumb, ArcGIS uses the default bandwidth that is determined as the minimum dimension (x or y) of the extent of the point theme divided by 30. Bailey & Gatrell (1995) suggested the following equation.

$$h_1 = 0.68(n)^{0.2}$$

In addition they introduced the adjusted equation depending on the size of the study area (*A*) as follows.

$$h_1 = 0.68(n)^{0.2}\sqrt{A}$$

However, results of these rules of thumb of bandwidth estimation do not take into account the spatial distribution of points (Williamson, McLafferty, Goldsmith, Mollenkopf, & P, 1999). Alternatively, it is suggested to use the average k-th nearest neighbor distance among points, in which small *k* values result in a small bandwidth producing a spiky map whereas larger *k* values result in a larger

bandwidth and smoother density map (Williamson, McLafferty, Goldsmith, Mollenkopf, & P, 1999). Nevertheless, Chainey and Ratcliffe (2005) pointed out by quoting Bailey and Gatrell (1995, pp. 86-87) that "the value of kernel estimation is that one can experiment with different values [of the bandwidth], exploring the surface, … using different degree of smoothing in order to look at variation in [the surface] at different scales."

STKDE calculates the density distribution from three-dimensional points; however, STPs are continuous features in a space-time cube described by discrete points. Some modifications to the scheme are required to handle STPs. This study proposes two approaches to apply STKDE for the density visualization of mobile objects. The first approach is to use three-dimensional vertices of STPs as an input point dataset. Because each segment, $S_{(k)}$, possesses scalar values of motion descriptors, $M_{(k)}$, they can be assigned to each vertex by averaging the values of two adjacent segments. Start and end vertices (i.e., edge points) of STP have only one adjacent segment; therefore, the values of the adjacent segment is directly assigned. Now, each vertex of a STP has motion descriptors, $Mp_{(j)}$. As an alternative way to represent motion descriptors, each STP can create a set of new vertices that is a medium coordinate of each segment, $Pm_{(k)}$. Then the value of motion descriptors of the corresponding segment can be directly assigned to $Pm_{(k)}$ (Figure 11). In these approaches, it is assumed that the sampling frequency of dataset is regularly fixed in time; otherwise the density distribution will be distorted. When sampling frequency is varied, it is necessary to resample by a regular time interval. In these approaches, it is important to mention that the value

of the point density distribution is affected by the sampling frequency of the dataset.

While the first approach uses points of segments by taking either vertices or middle points, the second approach uses voxel grids that partition a 3D space-time aquarium (Figure 12). Similar to a pixel for two-dimensional space, a voxel is a three-dimensional cell, which may contain several polylines. The center point of a voxel, $Pv_{(i)}$, assigned with average values of motion descriptors is then used to calculate STKDE. Furthermore, a scalar value of line density is added to $Pv_{(i)}$ by calculating the sum of line distances in each voxel divided by volume of the voxel. Whereas the first approach uses scalar values of motion descriptors of one segment, the second approach considers polylines in each voxel so that it can support measurements such as a straightness index and circular dispersion in addition to average values of basic motion descriptors.

To calculate motion descriptors in each voxel, the toolkit uses an algorithm for detecting an intersection point between a segment and a plane described by Ericson (2005). In Figure 13, let a plane P be given by $(n \bullet X)=d$ and a segment AB by the parametric equation as follows (Ericson, 2005).

$$S_{(t)} = A + t(B - A) \quad for \quad 0 \leq t \leq 1$$

The t value of intersection of the segment with the plane is obtained by substituting the parametric equation for X in the plane equation and solving for t as follows (Ericson, 2005).

$$n \cdot \big(A + t(B - A)\big) = d_{(t)}$$

$$t = (d - n \cdot A)/(n \cdot (B - A))$$

Now, *t* can be inserted into the parametric equation for the segment to find the actual intersection point *Q* (Ericson, 2005).

$$Q = A + [(d - n \cdot A)/(n \cdot (B - A))](B - A)$$

By applying the algorithm to each segment and each plane of voxel, intersection points will be detected. Now a voxel grid, *VG*, covering an entire 3D space-time aquarium, is composed of a set of voxel, $V_{(n)}$.

$$VG \in V_n$$

$$n = \frac{B_{vg_{(x)}} * B_{vg_{(y)}} * B_{vg_{(t)}}}{G_{s_{(x)}} * G_{s_{(y)}} * G_{s_{(t)}}}$$

where $Evg_{(x)}$, $Evg_{(y)}$, and $Evg_{(t)}$ are extents on each axis, $Gs_{(x)}$, $Gs_{(y)}$, $Gs_{(t)}$, are user choice of voxel size on each axis, and n is the total number of voxel.

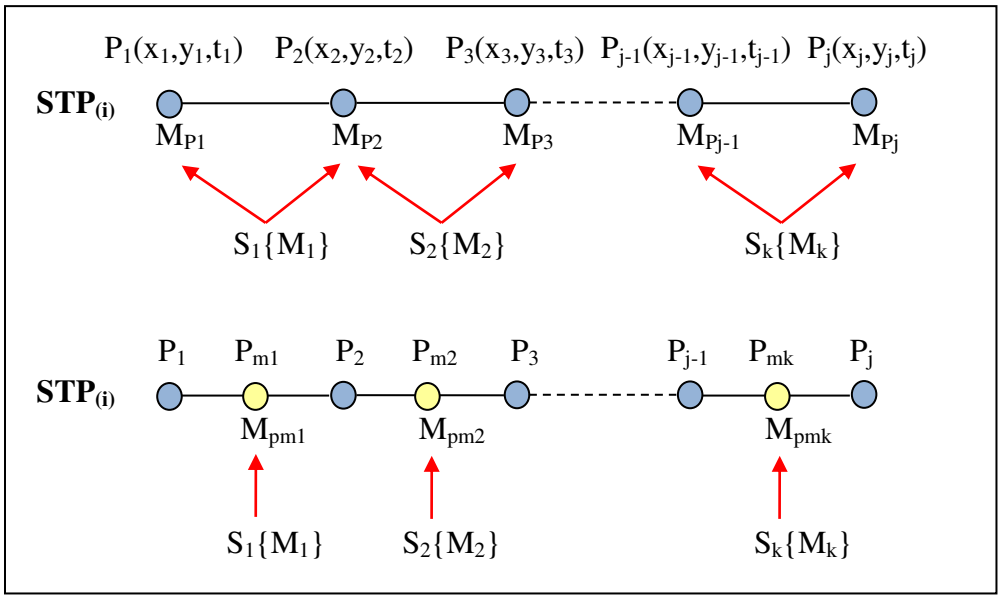

Figure 11. Image of assigning motion descriptors to vertices in a STP.

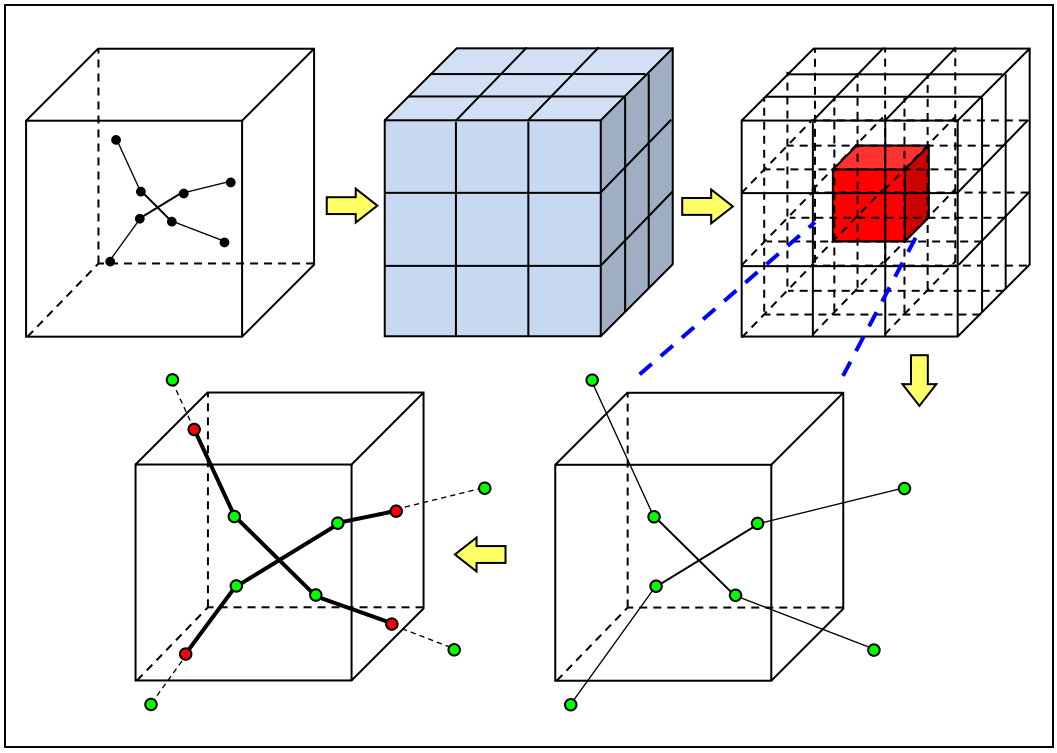

Figure 12. Partitioning a space-time aquarium and averaging motion descriptors.

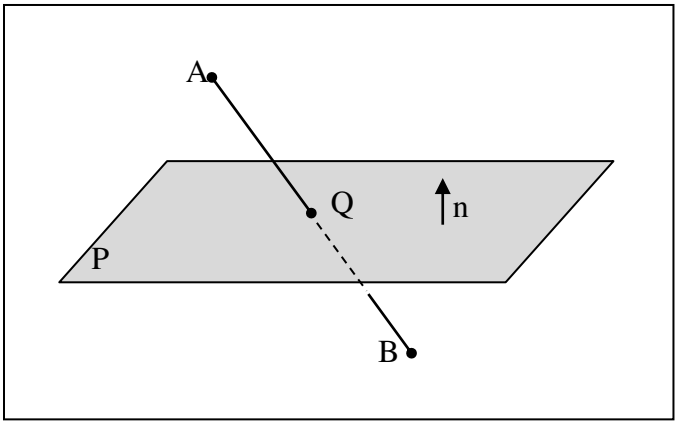

Figure 13. Intersection between a segment and a plane.

4.3.2.2 Volume rendering

STKDE generates scalar fields in a 3D space-time aquarium, where any point has 4-dimensional scalars (x,y,t,f) – two dimensions for geographical space (x,y), one for time (t), and one for the value of density estimation (f). To visualize the result of STKDE, Brunsdon, Corcoran, & Higgs (2007) employed an isosurface approach that is a two-dimensional surface embedded in three-dimensional space which joins together points having the same value (l) of a function (f) applied to the three arguments represented by the point.

$$\hat{f}(x, y, t) = l$$

The isosurface approach, however, encounters difficulty in visualizing multiple isosurfaces because isosurfaces of one value surround other isosurfaces in a three-dimensional space. Even though it is possible to apply translucence colors for outer isosurfaces, the approach does not really work at all for more than two isosurfaces (Brunsdon, Corcoran, & Higgs, 2007).

Volume rendering techniques are an alternative approach to visualize the result of STKDE as a 3D volumetric data in a single 2D image. It was first proposed by Levoy (1988) to visualize computed tomography (CT) data for medical imaging, and over the years many techniques have been developed for improving computation efficiency and visualization quality (Kaufman & Mueller, 2005). The process of Levoy's (1988) original approach of direct volume rendering includes shading, classification, ray-casting, and composing. Shading assigns a color to each voxel, while classification assigns opacity to each voxel. From the observer eye-point, two rays are then cast into voxel arrays for color and

opacity. Resampling of colors and opacities is computed at evenly spaced locations along the rays using trilinear interpolation. Resampled colors and opacities are then merged with each other and with the background by composing in back-to-front to yield a single color to determine pixel information for the output 2D image. Nakaya and Yano (2008; 2010) first applied STKDE and the volume rendering technique together to investigate space-time sequence of crime clusters/hotspots in Kyoto City, Japan.

For my work, the interactive approach of volume rendering was achieved using an open source visualization software, ParaView (Henderson, 2007). The ParaView system uses a Visualization Toolkit (VTK) data format, and the space-time exploration toolkit implements an output function that generates a VTK file of results from STKDE.

4.3.3 Space-Time Analysis Toolkit

I developed the main component of the Space-Time Analysis toolkit in the Microsoft .Net Framework using Visual C#. Data was stored in MySQL server. The tool supports manipulating data tables in the database server through Graphic User Interface (GUI) tools which I designed to cluster common tasks in hierarchical menus (Edit, Analysis). GUI provides the efficiency and ease of use for tools implemented in the toolkit.

Figure 14 shows the GUI of main display of the Space-Time Analysis toolkit and Figure 15 displays the GUI of the Space-Time Kernel Density Estimation Tool

with the image of voxel grids, points of vertices of STPs, and intersection points between STPs and voxel grids illustrated.

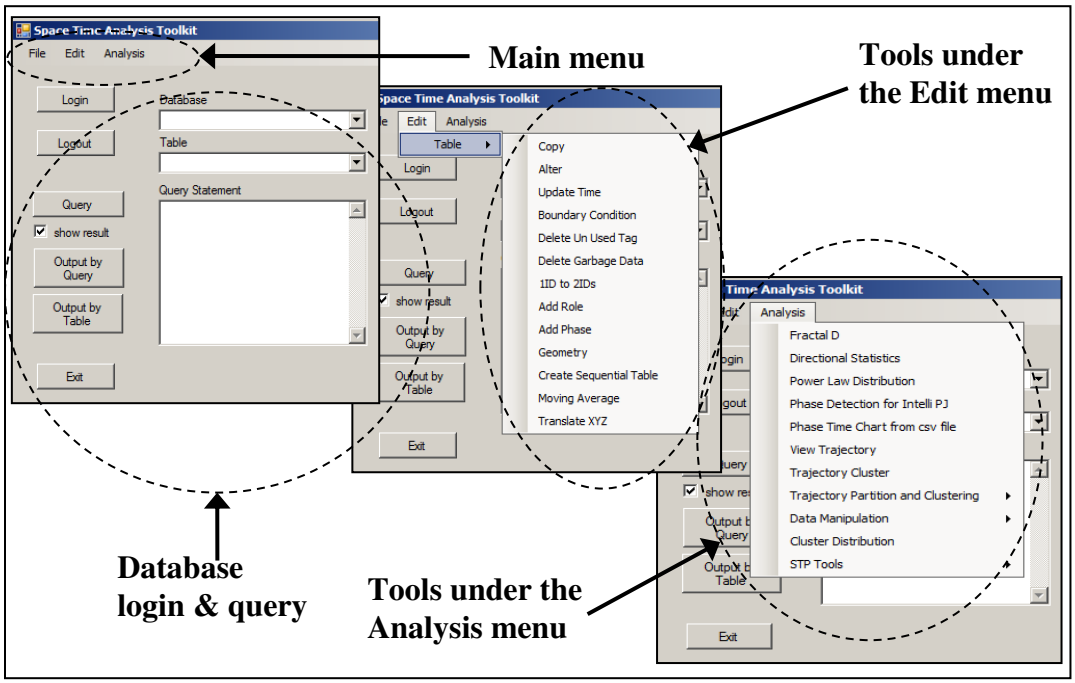

Figure 14. GUI of main display of the Space-Time Analysis toolkit.

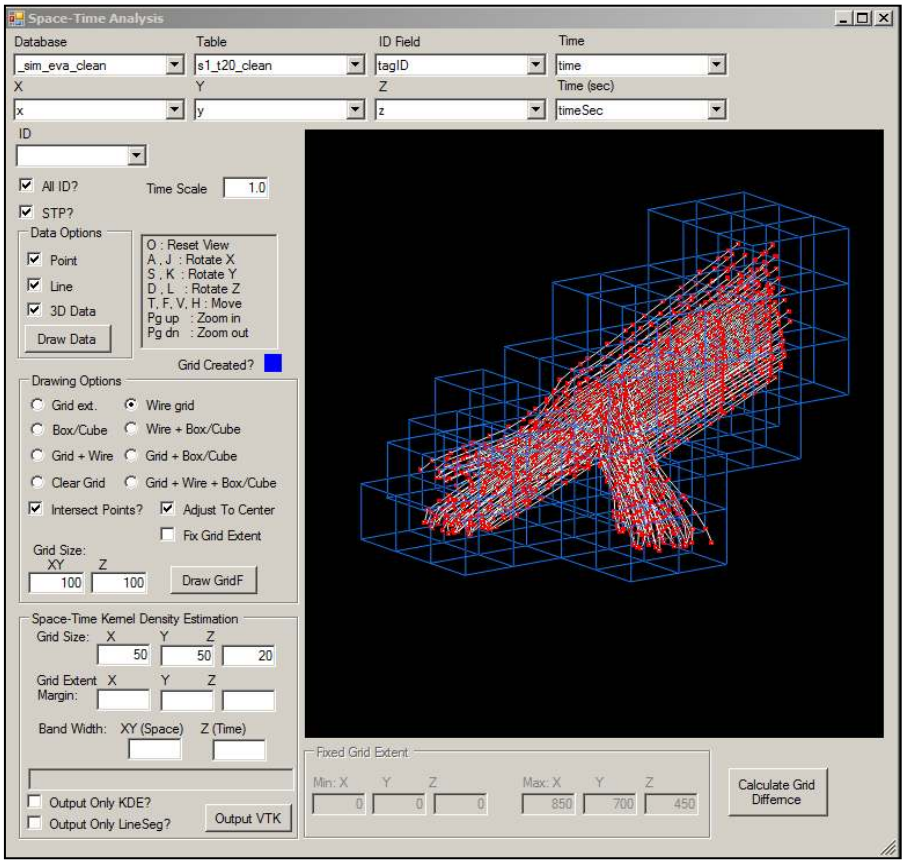

Figure 15. GUI of the Space-Time Kernel Density Estimation Tool

4.4 Case Study - Pedestrian Evacuation Dynamics

In the following section I will demonstrate how the toolkit can be used to illustrate, analyze, and visualize patterns and attributes of multiple STPs, and how the toolkit can be applied in a practical example. I focus on pedestrian evacuation dynamics, which I will later return to in more detail when describing agent-based modeling in Chapter 6.

4.4.1 Dataset

As a case study to examine the framework and toolkit developed, data regarding pedestrian evacuation dynamics is analyzed. The data is generated by an ABM based on the social force model developed by Helbing and Molnár (1995). The social force model is based on assumptions that a mixture of socio-psychological and physical forces influencing the behavior in a crowd (Helbing, Farkas, & Vicsek, 2000). In its simplest form, there are three forces formulated as follows.

$$m_i \frac{dv_i}{dt} = m_i \frac{v_i^0(t)e_i^0(t) - v_i(t)}{\tau_i} + \sum_{f(\neq \mathrm{i})} f_{ij} + \sum_{w} f_{iw}$$

The first force is a driving force toward a desired destination described by a pedestrian *i* of mass m_i, of desired velocity v_i^0, of desired direction e_i^0, and of actual velocity v_i with a characteristic time (acceleration time) τ_i. The second force is a repulsive force, $\sum_{j(\neq i)} f_{ij}$, describing the interaction effects with other agents *j* ($j \neq i$), and the third force is a repulsive force, $\sum_w f_{iw}$, to avoid walls and obstacles.

Pedestrians in this basic form of the social force model walk unidirectionally, i.e., each pedestrian travels between an origin and a destination. This is too simplistic, so to overcome the deficiency, the idea of multiple waypoints is implemented. In the algorithm, each pedestrian (*i*) owns a sequenced list of waypoints and walks toward the first waypoint in the list. When he reaches at the waypoint within a certain buffer zone described by a two-dimensional vector *bZ*(*bx*, *by*), the waypoint is removed from the list and the pedestrian walks toward the first waypoint in the new list until reaching the final destination.

In this study, pedestrian evacuation dynamics on a four-way intersection was simulated using the social force model. In the simulation, pedestrians evacuate from North, West, and South corridors to an East exit. The idea of multiple destinations/waypoints is implemented in the model so that pedestrians from North and South corridors are able to make a turn to evacuate to an East exit. The spatial extent of the model was set to 800 in width and 700 in height in the simulation unit length, and one unit length corresponds to 1/30 meters (area width=26.7m, area height=23.3m, corridor width & height=5.0m). A pedestrian is represented as a circle with the radius equals 10 (0.33m). Pedestrian's desired velocity, v_i^0, is approximately Gaussian distribution with a mean value of 1.3 m/s, which represents pedestrian walks in normal situation (Helbing, Buzna, Johansson, & Werner, 2005), and a standard deviation of 0.1 m/s. To determine waypoints for pedestrians evacuate from North and South corridors, waypoint zones (size: width=10, height=5) were manually introduced and each of these pedestrians randomly picks one waypoint in the zone. For these pedestrians, the x-coordinate of the final destination is the East boundary of the simulation area and y-coordinate is determined by adding a random perturbation value from the y-coordinate of the waypoint. The destination point for pedestrians evacuating from the West corridor is set to the East boundary for x-coordinate and the center of the corridor for y-coordinate. 40 pedestrians are randomly distributed in three starting zone (Total pedestrians = 120). For each pedestrian, three-dimensional points (x, y, t) are sampled every 1 second (every 100 frames) to create trajectory data. Figure 16 shows snapshots of the simulation at simulation time of 46, 347, 620,

and 1606 respectively (time unit: frame). In order to analyze data of simulated pedestrian evacuation dynamics, locations (x,y) of pedestrians and corresponding time stamps were output at every one second (= 100 frames). Figure 17 illustrates trajectories of pedestrian evacuation dynamics created from the output data.

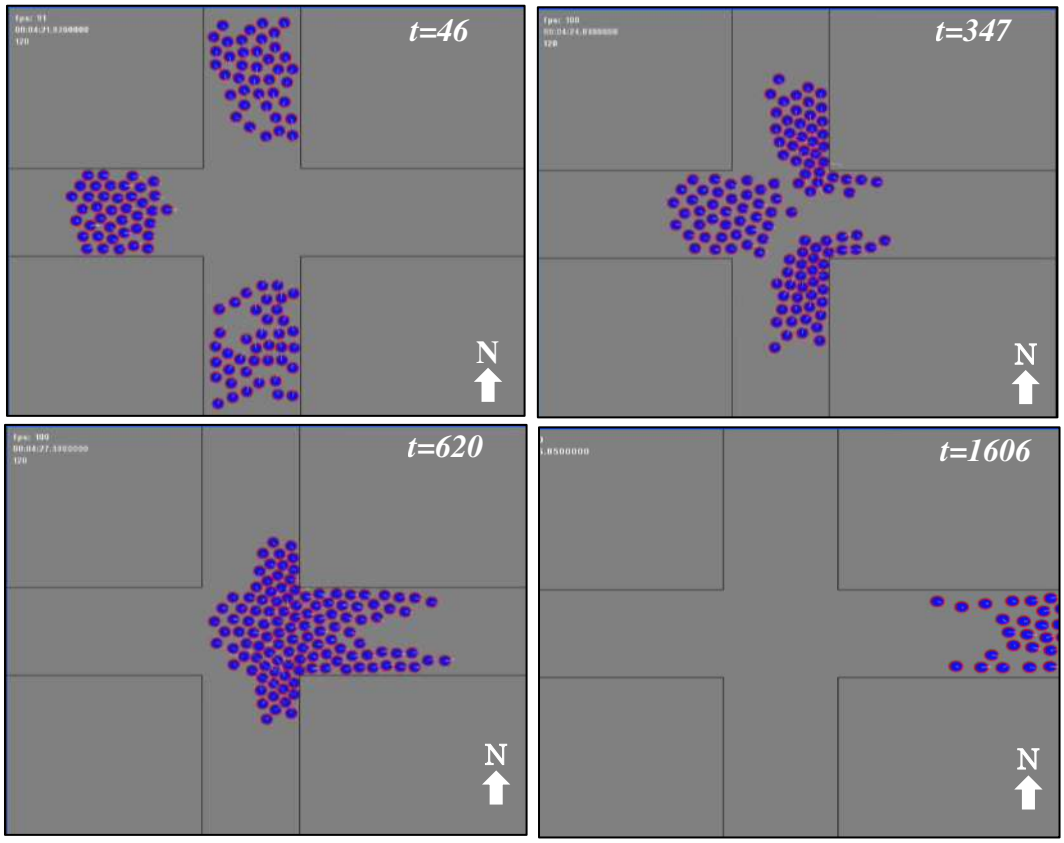

Figure 16. Simulation snapshots of pedestrian evacuation dynamics (time unit: frame).

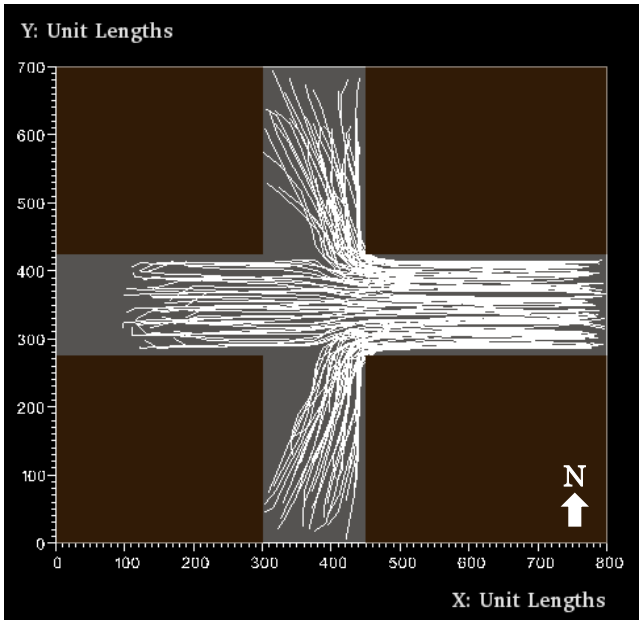

Figure 17. Trajectories of pedestrian evacuation dynamics.

4.4.2 Results of Quantitative and Qualitative Analysis of Crowd Evacuation Dynamics

Table 3 shows the result of quantifying trajectories as a whole and segments of trajectories by motion descriptors. Table 4 lists the correlation matrix of trajectories' motion descriptors. Figure 18 to Figure 21 illustrate two-dimensional maps of trajectories and each trajectory is colored by corresponding values of motion descriptors. As shown in Table 4 and Figure 18 to Figure 21, obvious high correlations are clearly visible; for example, negative correlation between evacuation time and average velocity; positive correlation among motion descriptors describing sinuosity of path including straightness index, fractal dimension, and circular dispersion. Figure 22 shows that the travel length of each

segment in the social force model forms a bell curve shape and does not follow the power-law. This is reasonable to explain pedestrian motion behaviors because the acceleration of human body motion is governed by greater mass and inertia and thus the speed distribution has a good agreement with normal distribution (Henderson, 1971). Figure 23 illustrates two-dimensional maps of trajectories colored by average velocity and acceleration of segments respectively, and Figure 24 visualizes corresponding two-dimensional maps of kernel density estimate (output grid size: 25 × 25 (unit length), bandwidth of KDE (h_1): 50 (unit length)). These show that the intersection where three groups of pedestrians meet is the bottleneck of evacuation, but they lack temporal information.

Table 3 Motion descriptors of trajectories (Trajectory: n=120, Segments: n=1674).

		Mean	SD	Min	Max
Trajectory	*Evacuation Time (sec)*	13.91	2.82	7.96	20.96
	Average Velocity (unit lengths / sec)	41.97	4.66	29.14	52.39
	Path Length (unit lengths)	572.29	67.61	398.81	683.89
	Straight Length (unit lengths)	506.48	98.33	352.55	680.70
	Straightness Index	0.8795	0.0885	0.7286	0.9996
	Fractal Dimension	1.0149	0.0120	1.0002	1.0649
	Circular Dispersion	0.1285	0.0916	0.0004	0.2772
Segment	*Average Velocity (unit lengths / sec)*	41.10	8.96	3.54	63.89
	Average Acceleration (unit velocity / sec)	1.20	7.36	-20.74	37.27

Table 4. Correlation matrix of trajectories' motion descriptors.

	Evac. Time	*Ave. Velocity*	*Path Length*	*St. Length*	*St. Index*	*Fractal D*	*Circ. Disp.*
Evac. Time	1						
Ave. Velocity	-0.8863	1					
Path Length	0.8367	-0.5085	1				
St. Length	0.6215	-0.2447	0.8890	1			
St. Index	0.2232	0.1053	0.5298	0.8580	1		
Fractal D	-0.4048	0.0828	-0.6753	-0.8364	-0.7996	1	
Circ. Disp.	-0.2447	-0.0859	-0.5456	-0.8641	-0.9956	0.8062	1

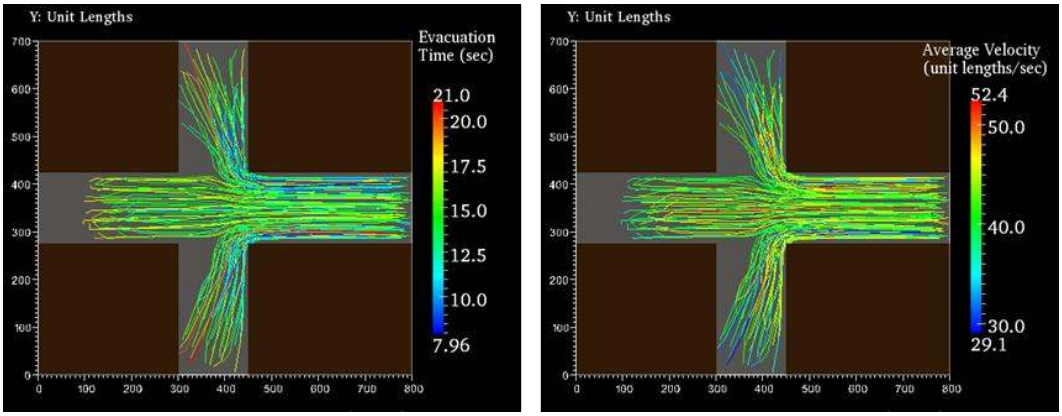

Figure 18. A 2D map of trajectories (left: evacuation time, right: average velocity).

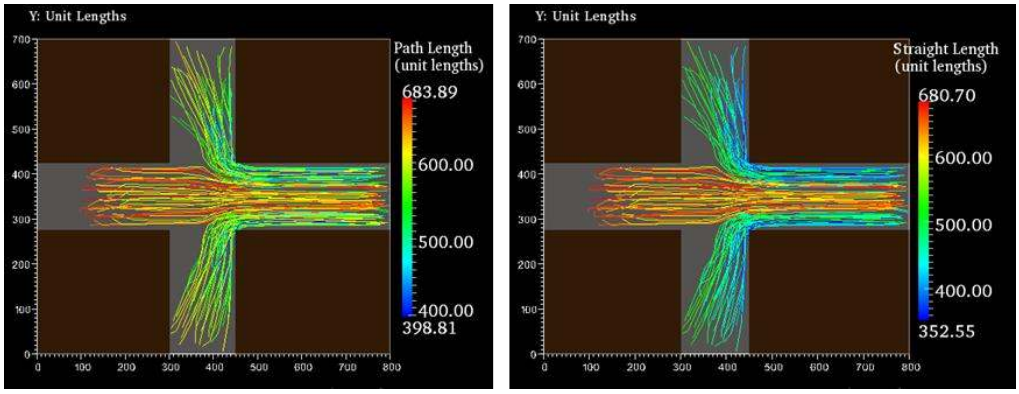

Figure 19. A 2D map of trajectories (left: path length, right: straight length).

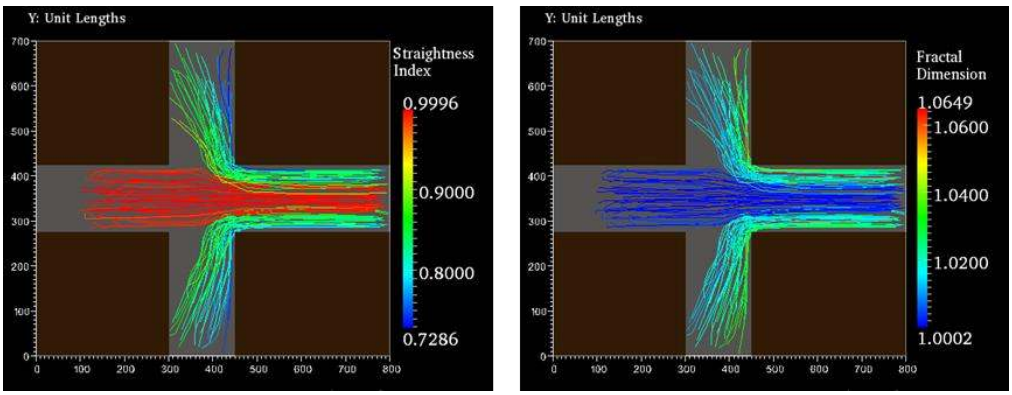

Figure 20. A 2D map of trajectories (left: straightness index, right: fractal dimension).

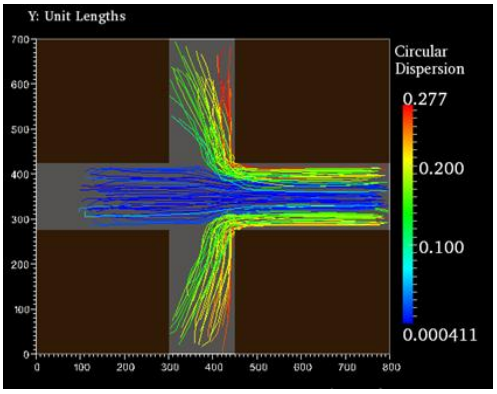

Figure 21. A 2D map of trajectories (circular dispersion).

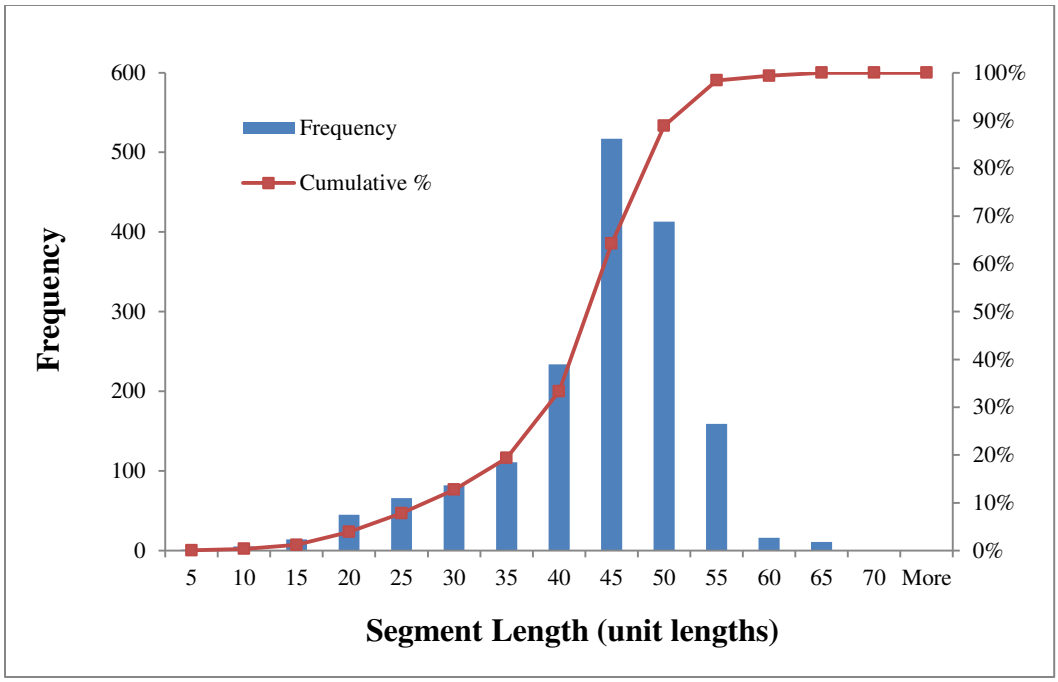

Figure 22. Histogram: A frequency distribution of length of segments (n=1674).

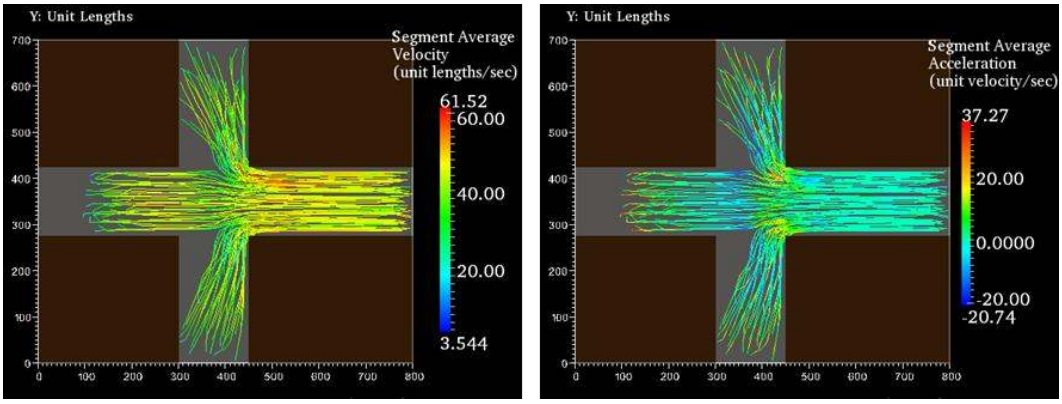

Figure 23. A 2D map of trajectories (left: average segment velocity, right: average segment acceleration).

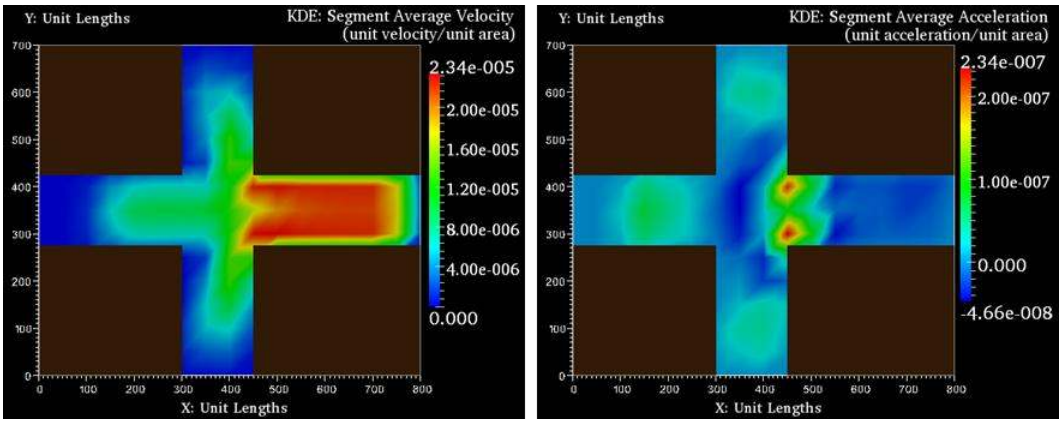

Figure 24. A 2D KDE map of trajectories (left: average segment velocity, right: average segment acceleration).

These visual/descriptive of trajectory motion descriptors can be further investigated under the framework of time geography to reveal spatio-temporal characteristics of pedestrians. In the dataset, the order of spatial scale (3 digits in unit length) is larger than that of temporal scale (2 digits in second). To exaggerate the temporal effect of crowd behaviors, the value of the time attribute is multiplied by 20. Figure 25 and Figure 26 illustrate stream-tubed STPs, colored and enhanced tube radius by average velocity and average acceleration of segments respectively. Red and thick tubes denote higher values, while blue and thin tubes are lower values. These representations allow us to identify spatio-temporal patterns of movement behaviors such as how a bottleneck is created and diminished in space and time. However, these representations only show the surface of multiple STPs and much of the movement behaviors are hidden due to the occlusion effects created by multiple paths.

Figure 27 shows Space-Time volume density maps using the voxel grid averaging approach (output voxel grid size: 40×40 (unit length) ×20 (unit time), bandwidth of STKDE: h_1=80 (unit length), h_2=40 (unit time)). Space-time density maps using the volume rendering technique can better support visual representation of these details as hot/cold spots of crowd movement behaviors described by motion descriptors. The top image of Figure 27 shows high values of line density (unit: unit lengths × unit area^{-1} × unit time^{-1}) in the East corridor near the intersection indicating the evacuation bottleneck. The velocity density map (Figure 27: middle-left image), on the other hand, highlights smooth evacuation behaviors in space and time. Higher values of acceleration are noticeable before space-time hot spots of velocity are observed (Figure 27: middle-right image). The density distribution of straightness index is consistent with high values through space and time except the spot around the corners of the intersection (Figure 27: bottom-left image). This is because the size of voxel grid in space is 40×40 and most partitioned trajectories fall in the grid are directed path. High values of circular dispersion density are observed around the intersection as well as pedestrian starting locations.

In summary, the illustrations show pedestrian egress behaviors as a collective movement in space and time. With various motion descriptors, they highlighted the hallmark of egress dynamics, specifically when and where pedestrian congestions took place described by high line density, low velocity and acceleration, low straightness index, and high circular dispersion. Identifying spatio-temporal pattern and process of collective and detail motion behaviors is

useful for better facility design as well as decision makings of evacuation route planning and scheduling.

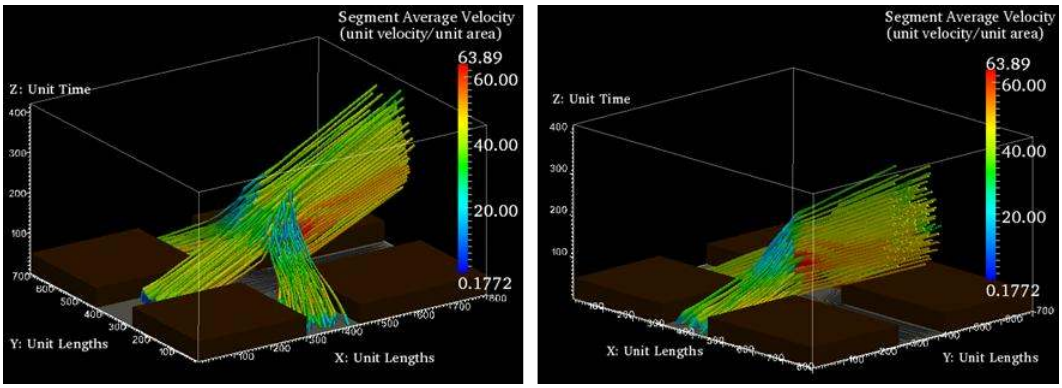

Figure 25. Stream-tubed STPs colored by average velocity of segments (left: a view from south west, right: a view from south east).

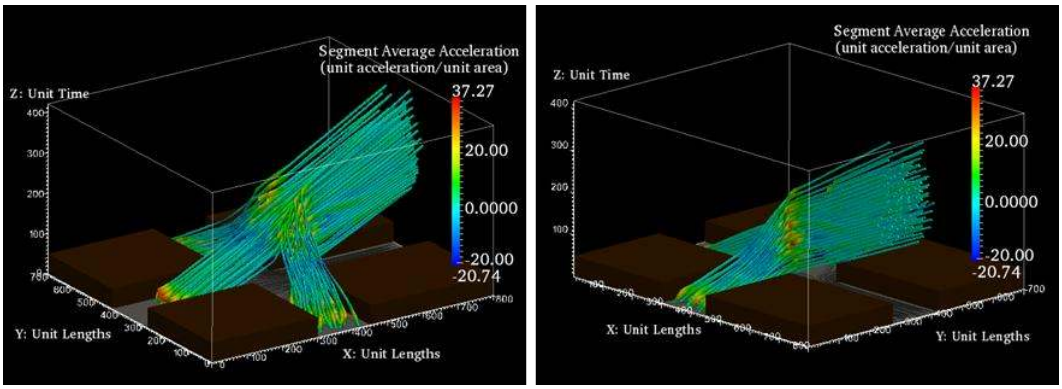

Figure 26. Stream-tubed STPs colored by average acceleration of segments (left: a view from south west, right: a view from south east).

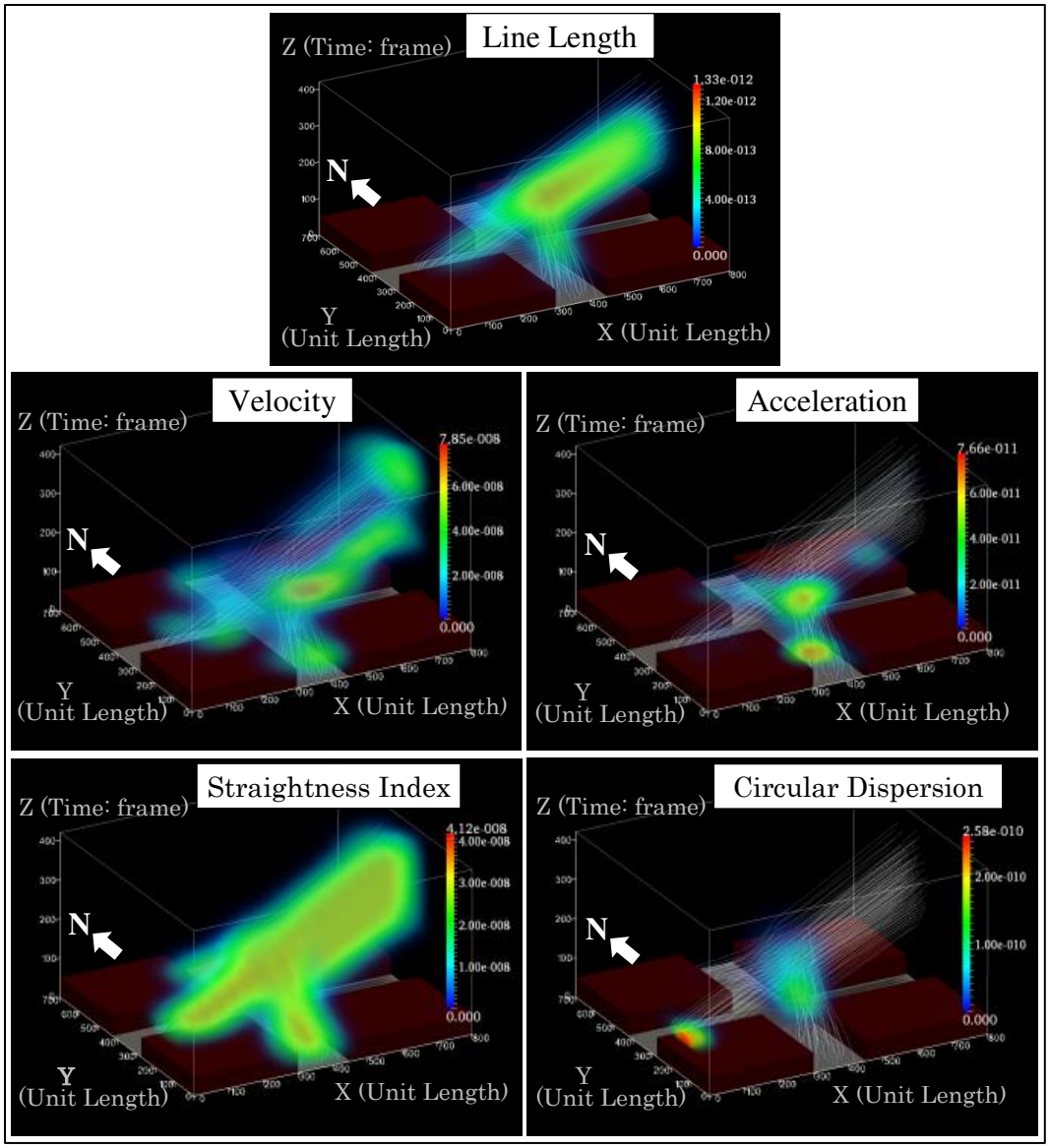

Figure 27. Space-Time volume density maps. The line density map (top) captures high crowd density in space and time, which is inversely related to the velocity density map (middle-left). Velocity (middle-left) and acceleration (middle-right) density maps describe human physical motion behavior, in which high acceleration is required to generate high velocity movement. Bottom images of straightness index and circular dispersion, both explaining path sinuosity, has an

inverse relationship. High directional disturbance is captured near the intersection corners as well as pedestrian's starting location.

4.5 Discussion and Conclusions

In this paper, I developed a novel analytical toolkit that describes and visualizes space-time motion behavior from a large dataset of multiple mobile objects. The toolkit allows us to explore and understand detail movement behaviors, their interactions, and collective behaviors through space and time. The results from the case study presented the functionality, capability, and effectiveness of the toolkit. A case study specifically analyzed pedestrian crowd evacuation dynamics and described behavioral pattern and process of crowd congestion; however, the toolkit can be applicable to wide variety of movement datasets that ubiquitously exist in nature.

This part of the work sought to develop an integrated spatio-temporal data exploration toolkit to represent spatio-temporal pattern and process of multiple mobile objects. The toolkit integrated both quantitative and qualitative representations of mobile objects. It incorporated the ability to calculate various motion descriptors and to capture movement characteristics of individual mobile objects as a whole trajectory and a collection of segments. The toolkit also implemented a qualitative visualization technique based around the concept of time geography using STKDE and volume rendering. It provides new insights for understandings of spatio-temporal behavioral pattern and process in large and complex data of mobile objects. The case study demonstrates that collective

movement behaviors of pedestrian crowd under evacuation scenario can be quantitatively and qualitatively described, even for massive amount of data and for complex scenarios with many interacting movements. The results capture and describe collective behavior of crowd congestion, an important feature of evacuation dynamics, in detail in space and time. Such results can be used for better facility design as well as decision makings of evacuation route planning and scheduling.

There are two considerations important in advancing the analytical power of the toolkit. First, the quantitative analysis implemented in the toolkit looks at movement behaviors by individual motion descriptors, which provides essential movement characteristics. However, motion behaviors can be better understood and meaningful when they are explained by multiple descriptors. For example, low velocity with large sinuous movements might describe wandering behaviors at a shopping mall, while high velocity with low sinuous movements might explain regular commuting behavior. To describe movement behaviors by multiple motion descriptors, it is useful to incorporate some aggregation techniques such as classification techniques that group similar movement behaviors.

Second, selection of voxel grid size and kernel bandwidth is an important issue. STKDE, with the voxel grid approach, partitions the space-time cube into regularly spaced voxels and summarizes motion behaviors in the voxels. Selecting too large size of voxel grids or kernel bandwidth may over smooth movement behaviors in space and time, while too small size may over localize movement

behaviors. This may introduce MAUP in the scale effect. Furthermore, this approach partitions trajectory by space and time but not by movement behaviors, which may be a potential source of MAUP of the zonal effect. Thus, finding optimal values for voxel grid size as well as bandwidths for STKDE is a research challenge, and sensitivity analysis on parameter selection will be a future research issue.

Chapter 5

TRAJECTORY DATA MINING: CLUSTERING, CONTEXT RECOGNITION, AND SPATIO-TEMPORAL VISUALIZATION

5.1 Overview

An alternative approach to contextualizing movement patterns, that could work in support of the analysis methods described in Chapter 4, is to use data mining to learn on movement data. By learning on trajectory data, spatial and temporal knowledge could be discovered in massive datasets.

Trajectory-based data mining is a very active research topic in the field of Knowledge Discovery in Databases (KDD) in response to the influx of mobile object data. Using a set of spatio-temporal sequences of mobile object data collected from various types of Location Aware Technologies (LATs) or generated by simulation models, trajectory data mining discovers spatio-temporal knowledge through exercises including pattern detection, clustering, classification, generalization, outlier detection, and visualization. Potential applications across various fields include, for example, vehicle and pedestrian traffic control (e.g., transportation management and facilities design); Location-Based Services (LBS) (e.g., navigation assistance and mobile advertising); weather forecasting (e.g., hurricane trajectory prediction and risk analysis); law enforcement (e.g., video surveillance for criminal activities); animal conservation (e.g., tracking at-risk animal populations); and logistics for goods and human.

Three major research challenges have been identified from previous works; 1) how to characterize and generalize massive trajectories to extract

interesting patterns; 2) how to explain behavioral contexts of trajectories by those extracted patterns; and 3) how to visualize extracted patterns to overview and compare patterns and trends in space and time.

For the second part of my work, I tackle above-mentioned challenges of trajectory data mining and context awareness of trajectory dataset by developing a trajectory data mining toolkit. In the first study, an integrated spatio-temporal data exploration toolkit was developed to better understand spatio-temporal pattern and process of multiple mobile objects. The toolkit explains motion behaviors by calculating basic motion descriptors (i.e., velocity, acceleration, orientation, length, and sinuosity), fractal dimension, directional distribution, and Lévy metrics. These descriptors individually provide essential movement characteristics of mobile objects; however, behavioral explanation by single descriptor is limited because real-world motion behaviors are rather complex. Therefore, motion behaviors and behavioral contexts can be better understood and meaningful when they are explained by multiple descriptors. In the second study, new functionalities are introduced to the toolkit developed in the first study. These include a trajectory data mining analysis scheme that employs trajectory partitioning and clustering algorithms to extract behavioral patterns of mobile objects using multiple motion descriptors as well as visual analysis to display extracted patterns and trends in space and time.

To examine the capability of the toolkit for extracting interesting patterns, explaining behavioral context, and visualizing extracted patterns, two movement datasets were analyzed. The first dataset is generated by purely mathematical

models so that their movement behaviors are known. Therefore, it is useful to examine how the proposed toolkit answers three research challenges. The dataset consists of mixed trajectories simulated by three random walk models; Brownian Motion (BM), Correlated Random Walk (CRW), and Lévy flight. As the second dataset, GPS tracks of real movement were used to test the data mining scheme on real-world data.

In summary, the results demonstrated that local behaviors of trajectory were well extracted to explain the global behavioral context from mixed trajectories of random walkers. Extracted local behaviors in the GPS dataset differentiated real movement activities during a day; however, the explanation power for global behavioral context recognition by local behaviors is not much improved from the recognition by global behaviors. These results indicate that the proposed trajectory data mining framework performs well on mixed behavioral datasets that are explicitly defined by mathematical expressions; however, when it applied to the real-world dataset to understand complex behaviors of human movements, the explanation power is limited.

5.2 Related Works

As the influx of data about mobile objects grows, there is increasing interest in performing data analysis over trajectory datasets to derive meaning from the data. Clustering has been popularly used to accomplish this, which is to group objects showing similar behavior and differentiate objects performing differently.

Focusing on the trajectory patterns of geometric shapes, Gaffney & Smyth (1999) proposed a model-based clustering algorithm by introducing a probabilistic mixture regression model and the Expectation-Maximization algorithm. The method was applied to identify the groups of similar trajectories of hand movements in video streams (Gaffney & Smyth, 1999) and extratropical cyclones (Gaffney, Robertson, Smyth, Camargo, & Ghil, 2006). In the approach, a trajectory is considered as a whole; however, a trajectory may have a long and complicated path so that only some portions of trajectories exhibit a common behavior, but the behavior is not common over the entire trajectory (Lee, Han, & Whang, 2007). Trajectory partitioning and clustering is an alternate approach to divide a whole trajectory into sub-trajectories and to conduct clustering analysis over sub-trajectories to extract similar behavior at the sub-trajectory level. This approach enables us to extract local behavioral patterns of mobile objects rather than global patterns. Lee, Han, & Whang (2007) proposed the sub-trajectory partitioning and clustering algorithm, TRACLUS. In the algorithm a whole trajectory is optimally partitioned into sub-trajectories based on the MDL (Minimum Description Length) principle, and then partitioned sub-trajectories are grouped into clusters based on density-based clustering, DBSCAN (Density-Based Spatial Clustering of Applications with Noise) (Ester, Kriegel, Sander, & Xu, 1996), in which a cluster is defined as a maximum set of density-connected points. Applying TRACLUS to the hurricane track dataset and the animal movement dataset (Elk and Deer), sub-trajectory clusters, representative

trajectories of those clusters, and outliers of sub-trajectories were identified (Lee, Han, & Whang, 2007; Lee, Han, & Li, 2008).

While the approaches above focus on the geometrical shape of a trajectory dataset, it is important to take into account the semantics of trajectory for useful knowledge discovery in practice (Bogorny, Kuijpers, & Alvares, 2009). There are several studies that focus on behavioral context recognition from mobile object datasets. For example, researchers at MIT Media Lab collected the Reality Data Mining Dataset, which covers one hundred human subjects' data about location, communication and device usage behavior using smart phones over nine months (Eagle & Pentland, 2006). Eagle and Pentland (2006; 2009) represented the structure of behavioral contexts of individuals, such as staying at home, work, or elsewhere, described by the principal components of the dataset, termed eigenbehaviors. Patterson, et al. (2003) collected GPS data, which contains position and velocity information sampled at 2-10 second intervals over three months during outside activities. The dataset was then hand labeled with one of three transportation modes; foot, bus, or car (Patterson, Liao, Fox, & Kautz, 2003). The authors enriched the dataset by adding external knowledge about bus routes and stops, and obtained 84% accuracy to predict modes of transportation using particle filters (Patterson, Liao, Fox, & Kautz, 2003). The analysis of these approaches is based on not only trajectory information acquired from LATs (i.e., a sequence of locational information (x, y) and time stamps (t)) but also other information such as behavioral and environmental data. Availability of a context-rich dataset is a critical success factor for empirically based research; however,

such datasets are often not easily accessible due to the cost, security and privacy issues (Giannotti & Pedreschi, 2007).

Other approaches consider movement behaviors of mobile objects by calculating various motion descriptors from trajectory datasets. For example, a trajectory is described by characteristics such as travel length, beeline length, speed, acceleration, duration, sinuosity, and direction. Dodge, Weibel, and Forootan (2009) argued that different types of mobile objects, depending on the particular physics of their movement, to some degree exhibit different signatures of such movement descriptors. Thus, similarity and dissimilarity of behavioral patterns of mixed mobile objects can be explained by one or several motion descriptors. Laube, et al. (2007) introduced a methodology for lifeline context operators and standardisations, and explored the spatio-temporal behaviors of homing pigeons using the sinuosity, rate of change of trajectory sinuosity, navigational displacement, relation between distance to loft and flight sinuosity. Dodge, Weibel, and Forootan (2009) distinguished trajectories of different transportation modes (i.e., motorcycle, car, bicycle, pedestrian) by global and local motion descriptors (e.g., velocity, acceleration, turning angle, straightness index), applied Principal Component Analysis (PCA) to reduce dimensionality of descriptors, and classified data using a supervised learning technique of Support Vector Machine (SVM). Problems remain, however, because mobile objects are ubiquitous in physical nature as well as cyber space and plenty of potential applications exist, many data mining methodologies are currently being developed in a piecemeal/ad hoc fashion and have yet to migrate from research to

demonstrate convincing social and commercial benefits (Weibel, Sack, Sester, & Bitterlich, 2008). Cao, Mamoulis, and Cheung (2009) also claimed that it is necessary to develop a fundamental theory and systematic framework for modeling and analyzing trajectories of mobile objects.

Visualization of trajectory patterns is another research challenge in trajectory data mining. Simple visualization techniques of trajectory in a 2D map or a 3D space-time cube are constrained in representing patterns and trends of massive movement data due to the cluttering and overlapping of symbols; thus, it is necessary to apply some forms of data aggregation and generalization (Andrienko & Andrienko, 2011). Guo, Liu, and Jin (2010) proposed a graph-based partition method by incorporating the use of trajectory topological relationships to find spatial structures and general patterns of trajectories, and visualized in 2D trajectory density maps at several temporal snapshots. Andrienko and Andrienko (2011) introduced a trajectory aggregation technique by partitioning the space into compartments, transforming raw trajectory data into moves between the compartments, and aggregating the transformed moves with common origins and common destinations. Then the authors visualized the aggregate information of moves by means of a flow map at various spatial and temporal granularities. Shen and Ma (2008) visualized social-spatial-temporal patterns of mobile data by developing a toolkit, MobiVis. The tool incorporated heterogeneous network and semantic filtering techniques based on associated ontology graphs (Shen, Ma, & Eliassi-Rad, 2006), and the visualization technique of behavior rings that reveal periodical behavioral patterns of individuals and

groups. Willems, Wetering, & Wijk (2009) applied Kernel Density Estimation (KDE) to visualize movement patterns of seafaring vessels. They computed trajectory density at two spatial scales (large and small), and simultaneously displayed both densities by shading the large scale density with a height map of the accumulated densities.

Time geographical visualization is another approach to exploratory investigation of spatio-temporal patterns of mobile objects (Kwan, 2000a; Kapler & Wright, 2004; Yu & Shaw, 2008; Miller & Bridwell, 2009); however, it also suffers from difficulties in visualizing massively mobile objects (Kwan, 2000a; Shaw, Yu, & Bombom, 2008). To overcome the deficiency, Shaw, Yu, and Bombom (2008) proposed to create generalized space-time paths (GSTPs) by identifying representative locations to portray the spatial distribution patterns of individuals at specified time periods using k-means clustering, and connecting the representative locations according to their temporal sequence. The authors developed an exploratory toolkit and implemented the GSTP algorithm using commercial GIS software and successfully demonstrated the capability of time geography to exploratory analysis and geovisualization of spatio-temporal patterns and trends in mobile objects' datasets. However, current popular GIS software can only handle geographic data in 2D or 2.5D (i.e., single value of Z coordinate), but have difficulty in handling 3D data (i.e., multiple Z coordinates) and beyond (Abdul-Rahman & Pilouk, 2008). Thus, visual inspection has not been fully investigated, or resolved.

This second component of my research offers new insights into current challenges in trajectory data mining by developing a trajectory data mining framework and toolkit. The functionalities of the toolkit include partitioning and clustering trajectories to extract similar movement behaviors from massive trajectory dataset, reconstruct behavioral contexts of trajectories from extracted movement behavioral patterns, and visualize extracted information under the concept of time geography to exploratory analyze spatio-temporal patterns and trends in mobile objects.

This proves to be useful because of following reasons. First, the trajectory data mining framework allow us to explore massive and complex spatio-temporal datasets of mobile objects and to extract hidden patters, trends, behavioral contexts, and useful information and knowledge. Second, human activities are typically composed of multiple movement behaviors across scales in space and time. For example, a commuting activity for urban residents can be described by motion behaviors such as direct walking, running, and waiting for a train, while a shopping activity at a mall may involve wandering and staying at multiple places. Therefore, to describe human activities and behavioral contexts from trajectory datasets, it is better to capture local motion behaviors rather than to use aggregated motion behaviors because they can easily loose behavioral variations. The proposed trajectory partitioning and clustering scheme naturally fits the concept because it decomposes a trajectory into a set of sub-trajectories that have similar motion characteristics, and classifies and extracts key local motion behaviors that can be used to explain human activities. Third, advanced

visualization techniques greatly help data mining exercises because the human visual system is extremely effective at recognizing patterns trends, and anomalies (Miller & Han, 2009). In particular, this study employs the concept of time geography that is useful to visualize and explore how human activities regarding to motion behaviors are distributed in space and time.

To evaluate the capability of the toolkit, two movement datasets were analyzed; 1) mixed movements generated by three different random walk models, Brownian Motion, Correlated Random Walk, and Lévy flight; and 2) recorded, real-world human movements in urban space collected by a GPS device.

5.3 Methodology

This study assumes that global behaviors of mobile objects (e.g., shopping, commuting, working, and traveling) in space and time are composed of multiple local behaviors (e.g., walking, running, turning, queuing, driving, and staying). The aim of developing a framework and a toolkit of trajectory data mining is to identify local behaviors of movement patterns from raw trajectory datasets. The contexts of global behaviors of mobile objects are then explained by the composition of extracted local behaviors. The proposed methodological framework includes three steps; trajectory partitioning, trajectory clustering, and evaluation of trajectory clustering.

- Step1: Trajectory partitioning
    - TRACLUS with MDL

    - Distance-Threshold
- Step 2: Trajectory clustering
    - Quantification of sub-trajectory
    - Principal Component Analysis
    - K-means cluster analysis
        - Gap statistics for searching optimal K
- Step3: Evaluation of trajectory clustering
    - Behavioral recognition by decision tree
    - Visualization of trajectory cluster distribution

Trajectory partitioning is the first process to partition a single trajectory into a set of sub-trajectories in order to extract local motion behaviors in the trajectory. Two approaches were implemented in the toolkit. The TRACLUS with MDL approach partitions a trajectory by finding a significant change in geometry, while the Distance-Threshold approach uses a distance value to find staying activities in a dataset and then partitions a trajectory at the staying points.

Using the sub-trajectory dataset, the second process is trajectory clustering in order to group sub-trajectories with similar motion characteristics. There are three sub-steps. The first sub-step calculates motion descriptors for each sub-trajectory and obtains a multi-dimensional vector. As the second sub-step, PCA is used to reduce the dimensionality of the sub-trajectory dataset because the dataset described by the multi-dimensional vector consists of interrelated motion variables (e.g., segment length and duration). The third sub-step is an

unsupervised cluster analysis to classify sub-trajectories for extracting local movement behaviors using the K-means clustering algorithm.

The third process is to evaluate identified behavioral clusters of trajectories using two approaches; a supervised classification of decision tree and a visual investigation of trajectory cluster distribution based on Space-Time Paths (STPs) and Space-Time Kernel Density Estimation (STKDE) utilizing a volume rendering technique.

5.3.1 Trajectory Partitioning

A set of trajectories, which can be generated by simulation or collected from LATs, is described as {Trajectory Set: $TR_{set}=TR_1,TR_2, TR_3, \ldots, TR_i$, where i denotes the number of mobile objects} (Figure 28). Each trajectory is composed of a sequence of 4-dimensional points $\{\{TR_i=p_1, p_2, p_3, \ldots, p_j$, where j denotes the number of points in the trajectory $i\}$, $\{p_j=x, y, z, t\}\}$. The trajectory partitioning process partitions an entire trajectory of an individual into trajectory partitions (sub-trajectories). By grouping trajectory partitions, the clustering process describes different human activities in relation to movement behaviors. There are two trajectory partitioning algorithms implemented in this study, TRACLUS (Lee, Han, & Whang, 2007) based on a MDL principle and a Distance-Threshold approach.

The first algorithm finds the points, called characteristic points (p_c), where the behavior of a trajectory changes rapidly. This approach essentially considers the directionality of a trajectory, which is particularly useful to extract behaviors

when mobile objects show a behavioral change accompanied by their directional change in movements such as hurricanes and animal seasonal migrations. Each characteristic point partitions a trajectory into trajectory partitions and each partition is represented by a set of line segments between two consecutive characteristic points (Lee, Han, & Whang, 2007). $\{TR_{i=}TRpar_{(1)}\{p_{c(1)}, p_{c(2)}\}, TRpar_{(1)}\{p_{c(2)}, p_{c(3)}\}, ..., TRpar_{(m)}\{p_{c(n-1)}, p_{c(n)}\}\}$, where m denotes the number of trajectory partitions and n denotes the number of characteristic points ($m=n-1$). The optimal partitioning of a trajectory is achieved by two contradictory properties: preciseness and conciseness. Preciseness refers to the minimization of the difference between a trajectory and a set of its trajectory partitions, whereas conciseness refers to the minimization of the number of trajectory partitions. The optimal trade-off between preciseness and conciseness is approximated based on the MDL principle (Lee, Han, & Whang, 2007; Nara, Izumi, Iseki, Suzuki, Nambu, & Sakurai, 2009).

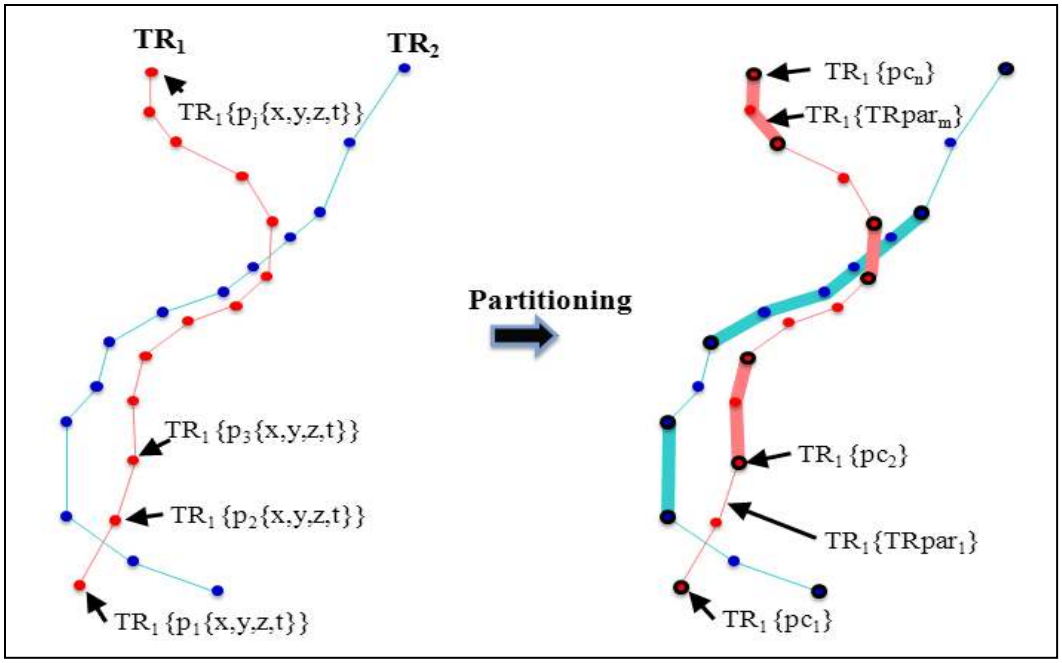

Figure 28. Trajectory partition (TRACLUS with the MDL approach).

The TRACLUS approach (Lee, Han, & Whang, 2007) employs a distance function that is composed of three kind of distances between two segments (*Li* (P_{i1}, P_{i2}), *Lj* (P_{j1}, P_{j2})); perpendicular distance ($d_{\perp}$), parallel distance ($d_{\parallel}$), and angle distance (d_{θ}). Figure 29 illustrates three components of the distance function. Projection points of P_{j1} and P_{j2} onto *Li* are P_{p1} and P_{p2} are shown. The Euclidean distances between P_{j1} and P_{p1} and between P_{j2} and P_{p2} are defined as $l_{\perp 1}$, $l_{\perp 2}$ respectively, and the perpendicular distance between *Li* and *Lj* is then defined by the Lehmer Mean of $l_{\perp 1}$ and $l_{\perp 2}$ with the order of 2 as follows.

$$d_{\perp}(L_i, L_j) = \frac{l_{\perp 1}^2 + l_{\perp 2}^2}{l_{\perp 1} + l_{\perp 2}}$$

The parallel distance between *Li* and *Lj* is defined as the minimum of the Euclidean distances of $l_{\parallel 1}$ and $l_{\parallel 2}$ as follows.

$$d_{\parallel}(L_i, L_j) = Min(l_{\parallel 1}, l_{\parallel 2})$$

The angle distance between *Li* and *Lj* described by the smaller intersecting angle between *Li* and *Lj*, θ, is defined as follows.

$$d_{\theta}(L_i, L_j) = \begin{cases} \|L_j\| \times \sin(\theta), & if \quad 0° \le \theta \le 90° \\ \|L_j\|, & if \quad 90° \le \theta \le 180° \end{cases}$$

The distance between two segments is finally defined as the sum of three distances.

$$d(L_i, L_j) = w_{\perp} d_{\perp}(L_i, L_j) + w_{\parallel} d_{\parallel}(L_i, L_j) + w_{\theta} d_{\theta}(L_i, L_j)$$

where, $w_{\parallel}$, $w_{\perp}$, and w_{θ} are the weights of each three distances respectively, and they are set equally to 1.0 as default values.

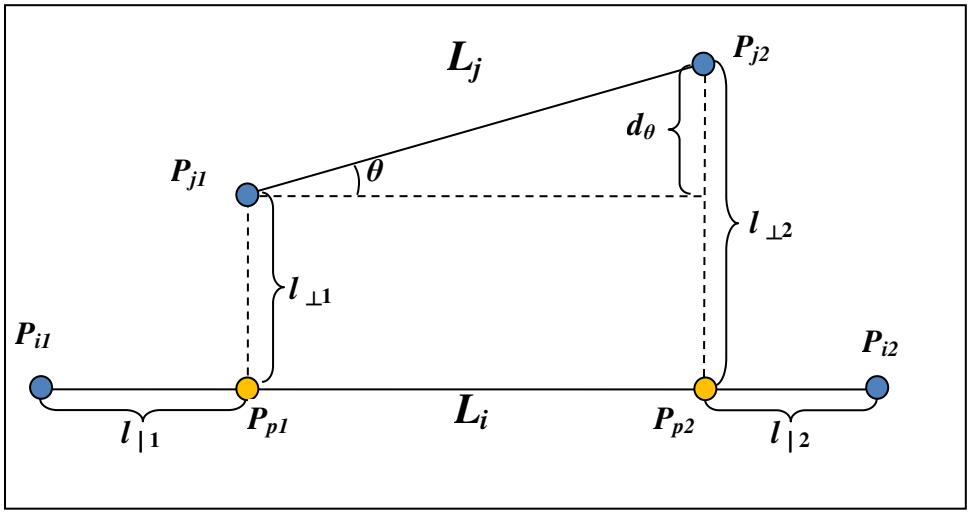

Figure 29. Three components of the distance function in TRACLUS. Adapted from Lee et al. (2007).

Using the distance function described above, the TRACLUS approach finds characteristic points that optimally partition a trajectory into trajectory partitions. The partitioning process is achieved by finding the optimal tradeoff between preciseness and conciseness based on the MDL principle.

In the principle, the MDL cost consists of two components; *L(H)* and *L(D/H)*. *L(H)* is the length of the description of the hypothesis *H*, and L(D/H) is the length of the description of the data *D*, and the best hypothesis *H* to explain *D* is the one that minimizes the sum of *L(H)* and *L(D/H)* (Grünwald, Myung, & Pitt, 2005). In the trajectory partition problem in TRACLUS algorithm, Lee, et al. (2007) defined that a set of trajectory partitions corresponds to *H* and a trajectory corresponds to *D*. They further defined the lengths of the hypothesis and the data as *L(H)* and *L(D/H)* respectively and these are mathematically defined as follows.

$$L(H) = c + \sum_{j=1}^{par_i - 1} \log_2 (len(p_{cj} p_{cj+1}))$$

$$L(D \mid H) = \sum_{j=1}^{par_i - 1} \sum_{k=c_j}^{c_{j+1} - 1} \left\{ \log_2 (d_\perp (p_{cj} p_{cj+1}, p_k p_{k+1})) + \log_2 (d_\theta (p_{cj} p_{cj+1}, p_k p_{k+1})) \right\}$$

where, *L(H)* measures the degree of consiseness calculated by the sum of logarithms of the two-dimensional Euclidean distance between two consecutive characteristic points in a trajectory. *L(D/H)* measures the degree of preciseness calculated by the sum of logarithms of the distances between a segment of a trajectory partition (p_{cj}, p_{cj+1}) and each line segment (p_k, p_{k+1}) residing in the traejctory partition. Thus, finding the optimal trajectory partitioning is obtained by finding the best hypothesis using the MDL principle (i.e., minimizes the sum of

L(H) and *L(D/H)*). The distance function is applied to calculate distance in the above equation; however, parallel distance is not considred because a trajectory enclosed its trajectory partition. *c* is the small constant for adjusting the partitioning criteria to suppress partitioning at the cost of preciseness; thus it increases the length of trajectory partitions.

Figure 30 illustrates the formation of the MDL cost.

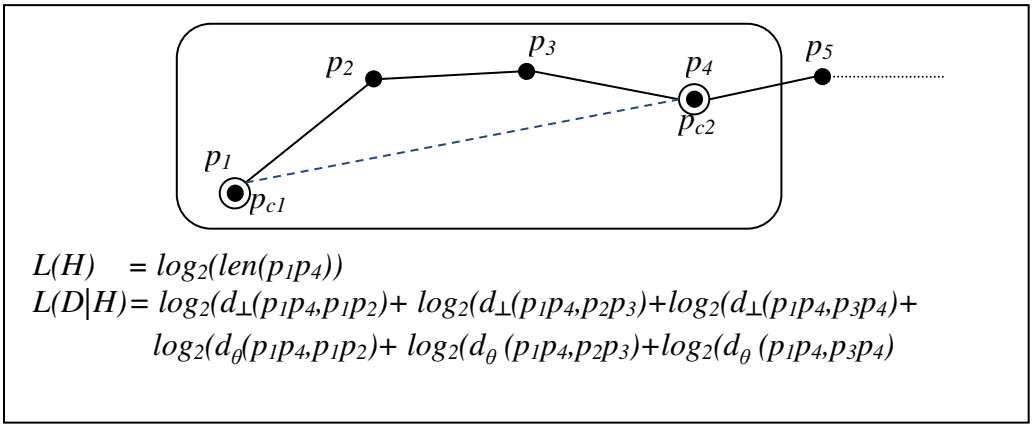

Figure 30. Formation of the MDL cost. Adapted from Lee, et al. (2007).

The optimal partitioning is to minimize the MDL cost, *L(H) + L(D/H)*; however, the cost of finding the global optimal partitioing is prohibitive because it is necessary to consider every subset of the points in a trajectory (Lee, Han, & Whang, 2007). In order to approximate the solution, Lee, et al. (2007) defined two MDL costs, $MDL_{par}(p_i, p_j)$ and $MDL_{nopar}(p_i, p_j)$. $MDL_{par}(p_i, p_j)$ is defined as the MDL cost of a trajectory between p_i and p_j (i < j) where there are only two characteristic points (p_i, p_j). $MDL_{nopar}(p_i, p_j)$ is defined as the MDL cost with no characteristic point between p_i, and p_j (i.e., preserving the original trajectory).

Then the approximate solution to minimize the MDL cost is obtained by the longest trajectory partition $p_i p_j$ that satisfies $MDL_{par}(p_i,p_k) \leq MDL_{nopar}(p_i,p_k)$ for every k such that $i \leq k \leq j$ (Lee, Han, & Whang, 2007).

The second approach is a Distance-Threshold based approach to partition a trajectory into sub-trajectories. This is a simple approach based on the assumption that in many situations movements of mobile objects involve with stopping/staying when the object changes its behavior. Such behaviors can be seen at multiple scales in human movements; for example, when a pedestrian decelerates and ultimately stops to make a sharp turn or to avoid collisions with other pedestrians; a commuter stays at home, walks to a bus stop, waits for a bus, takes a bus, and stays at its office to work; and a person may relocate and find a new home to stay associated with its life events.

Methodologically, partitioning a trajectory based on staying behavior can be simply achieved by introducing a Distance-Threshold (Th_d) (Figure 31). If a distance of each segment in a trajectory is less than Th_d, then the segment is assigned as *STAY* and a trajectory is partitioned by the segment. If consecutive segments are assigned to *STAY*, then those segments are considered as one sub-trajectory in order to differentiate staying behavior such as short stop or long stay. This grouping process introduces one problem, that is, a sub-trajectory is assigned as *STAY* if it is composed of multiple segments with each distance less than Th_d, but with same direction. This can be happen when a sub-trajectory describes very slow movement to one direction or when frequency of data sampling is fine. To avoid this mislabelling problem, a sub-trajectory composed of multiple segments

assigned as *STAY* is re-assigned as *MOVE* if the diameter of a minimum bounding circle of the sub-trajectory is greater than Th_d (Figure 32).

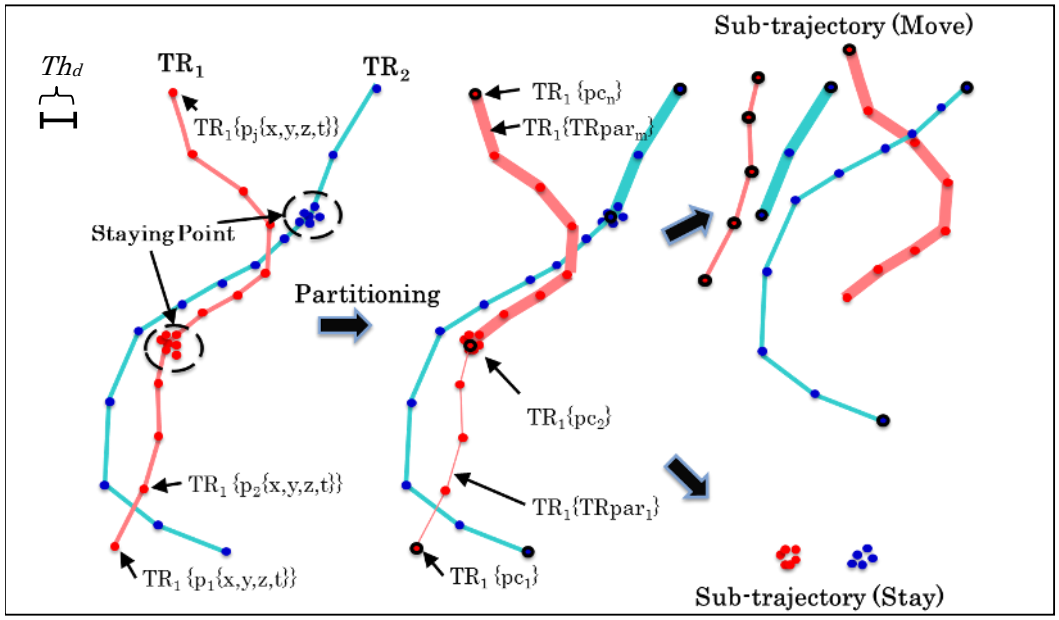

Figure 31. Trajectory partition (Distance-Threshold approach).

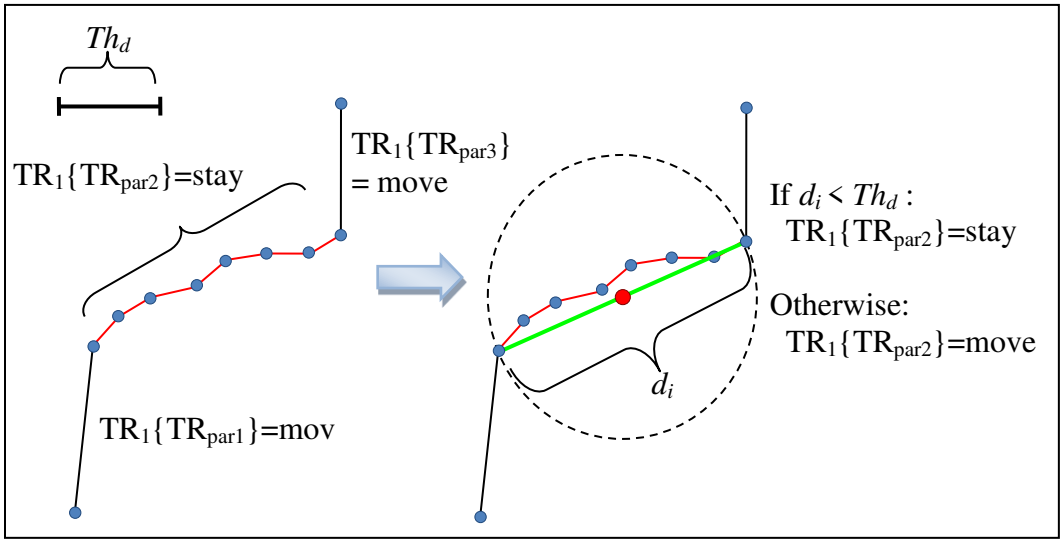

Figure 32. Labeling staying behavior on sub-trajectories.

5.3.2 Trajectory Clustering

For each trajectory partition ($TR_{par(i)}$), multi-dimensional vectors to characterize the partition trajectory are obtained. Each sub-trajectory is composed of m segments ($TR_{par(i)}=\{s_1\{ps_{11}, ps_{12}\}, s_2\{ps_{21}, ps_{22}\}, \ldots, s_m(ps_{m1}, ps_{m2})\}$, where each segment s_m is composed of two three-dimensional points (ps_{m1}:x,y,t, ps_{m2};x,y,t)). The vector values of a sub-trajectory include total duration (d_t), total horizontal distance (d_x), total vertical distance (d_y), total two-dimensional distance (d_{2D}), velocity vector on x-axis (v_x), velocity vector on y-axis (v_y), and velocity (v), horizontal distance between start and end nodes (d_{sx}), vertical distance between start and end nodes (d_{sy}), two-dimensional beeline distance between start and end nodes (d_{s2D}), area of minimum bounding box (mbb), and sum of cosine of turning angle between two consecutive segments (sct) as follows.

$$d_t = \sum_{i=1}^{m} ps_{m2}.t - ps_{m1}.t$$

$$d_x = \sum_{i=1}^{m} ps_{m2}.x - ps_{m1}.x$$

$$d_y = \sum_{i=1}^{m} ps_{m2}.y - ps_{m1}.y$$

$$d_{2D} = \sum_{i=1}^{m} \sqrt{\left(ps_{m2}.x - ps_{m1}.x\right)^2 + \left(ps_{m2}.y - ps_{m1}.y\right)^2}$$

$$v_x = \frac{d_x}{d_t}$$

$$v_y = \frac{d_y}{d_t}$$

$$v = \frac{d_{2D}}{d_t}$$

$$d_{sx} = ps_{m2}.x - ps_{11}.x$$

$$d_{sy} = ps_{m2}.y - ps_{11}.y$$

$$d_{s2D} = \sqrt{d_{sx}^2 + d_{sy}^2}$$

$$mbb = \left(\max\left(ps_m.x\right) - \min\left(ps_m.x\right)\right) \times \left(\max\left(ps_m.y\right) - \min\left(ps_m.y\right)\right)$$

$$sct = \sum_{i=2}^{m} \cos^{-1}\theta$$

$$\cos\theta = \frac{p_1 \bullet p_2}{d_{12}}$$

where,

$$p_1 = p\left((ps_{m-1,2}.x - ps_{m-1,1}.x), (ps_{m-1,2}.y - ps_{m-1,1}.y)\right),$$

$$p_2 = p\left((ps_{m,2}.x - ps_{m,1}.x), (ps_{m,2}.y - ps_{m,1}.y)\right),$$

$$d_{12} = \sqrt{\left(p_1.x \times p_1.x + p_1.y \times p_1.y\right) \times \left(p_2.x \times p_2.x + p_2.y \times p_2.y\right)}$$

When a cosine of turning angle equals 0, the turn made by a mobile object is 90°. A negative value of a cosine of turning angle represents a turn with more than 90°, while the value equals 1 with no turn. Thus, a large negative value of *sct* indicates that a path consists of many large turns, whereas a positive value indicates a path is composed of smooth turns. All of these vector values are then normalized with mean equals to 0 and variance equals to 1 (μ=0, σ=1) by the following equation.

$$y_i = \frac{(x_i - \mu)}{S} = \frac{\left(x_i - \bar{x}\right)}{\sqrt{\frac{1}{n}\sum_{j=0}^{n}\left(x_j - \bar{x}\right)^2}}$$

The normalization is an important procedure for PCA because units of variables in motion descriptors are different and variances in each variable may differ as well. If units of variables are different, Principal Components (PCs) retained by PCA will be different depending on the choice of units. In addition, if variances in variables largely differ, the result of PCA will be largely affected by variables with large variance; thus, it will be difficult to correctly interpolate the interrelationship among variables.

In order to reduce the dimensionality of multiple vectors of sub-trajectories, PCA may be used. PCA is a multivariate statistical technique to reduce the dimensionality of a dataset consisting of interrelated variables by finding a new set of variables, which is smaller than the original set of variables but still containing most of the information in the original dataset. This is achieved by transforming a set of original variables to a new set of variables, PCs, which are uncorrelated and ordered so that the first few retain the most of the variation present in all of the original variables (Jolliffe, 2002). The PCs are derived from the eigenvectors of the covariance or correlation matrix of the original variables, where a correlation matrix is used if each variable has different units of measure or the variances of variables differ large. Because vector variables of sub-trajectories are normalized in this study, the covariance matrix is used. Eigenvalues of PCs measure the amount of variation. To determine the number of PCs to retain, the Kaiser criterion (Kaiser, 1960) is introduced. The

criterion determines PCs to retain if the eigenvalue of PC is greater than 1 so that each PC explains at least as much variance as 1 observed variable. Next, PC scores of each sub-trajectory for each PC (Eigenvalue ≥ 1) are computed, and then they are used as a new input dataset for sub-trajectory clustering.

To classify sub-trajectories for extracting local movement behaviors, the K-means clustering algorithm is employed, where the input data is PC scores (eigenvalue >= 1) obtained for each sub-trajectory. As a non-hierarchical approach, the classical K-means clustering algorithm partitions *M* dataset in *N* dimensional variables into *k* groups (C_1,C_2, …, C_k) such that the total sum of squared Euclidean distances from each data point (x) to the centroid of the nearest group (c_i) in *N* dimensional space is locally minimized (MacQueen, 1967).

$$\sum_{i=1}^{K} \sum_{x \in C_i} \left\| x - c_i \right\|^2$$

In this study, the K-means clustering algorithm develped by Hartigan & Wong (1979) is used. To estimate the quality of clusters for determining the optimal value of *k* in K-means clustering automatically, clustering algorithms are run with different values of *k*, and the optimal value of *k* is selected by a predefined criterion such as Information Gain Ratio for Cluster (IGRC) (Yoshida, Shoda, & Motoda, 2006), Minimum Description Length (MDL) (Hansen & Yu, 2001), Bayes Information Criterion (BIC), Akaike Information Criterion (AIC), and Gap Statistics (Tibshirani, Walther, & Hastie, 2001), which is applied in this study.

In the gap statistics, W_k (k=1 to *k*) is defined as a within-cluster sum of squares of Euclidean distance around the cluster means measuring the

compactness of clusters. By generating *B* reference datasets of an appropriate null model, K-means clustering also gives the within-cluster sum of squares for each *B*, W_{kb} (k=1 to *k*, b=1 to *B*). In this study, reference datasets are set under uniform distribution over the *N* dimensional space of the observed data rage. The gap statistics estimates the optimal *k* value, $\hat{k}$ of by calculating the difference, $Gap_{(k)}$, of the expected value of log(W_{kb}) of null reference dataset and the log(W_k) of the observed dataset as follows.

$$Gap(k) = \frac{1}{B}\sum_{b=1}^{B} logW_{kb} - logW_k$$

$$\hat{k} = smallest\ k\ such\ that\ Gap(k) \geq Gap(k+1) - s(k+1)$$

where, $s(k) = \sqrt{1+\frac{1}{B}} * sd(k)$, $sd(k) = \sqrt{\frac{1}{B}\sum_b \{log(W_{kb}) - \bar{l}\}^2}$, $\bar{l} = \frac{1}{B}\sum_b log(W_{kb})$

5.3.3 Evaluation of Trajectory Clustering

5.3.3.1 Behavioral recognition by decision tree

To evaluate the quality of trajectory clustering, global behavioral contexts are reconstructed from extracted clusters of local movement behaviors. Contextual recognition of moving objects can be achieved by using supervised learning classification techniques such as decision tree induction (Quinlan, 1986), naïve Bayesian classification (Domingos & Pazzani, 1997), artificial neural networks (Bishop, 1995), maximum likelihood estimation (Fisher, 1922), and support vector machine (Cortes & Vapnik, 1995). This process can be done if the

reference data of behavioral context is available along with trajectory dataset. For example, in the case of behavioral recognition in human daily activity, reference data can be activity diary that may include daily activities (e.g., work day, day-off), transportation modes (e.g., walk, car, train), and major activities (e.g., working at office, staying at home, shopping, dining).

This study employs a decision tree classification algorithm, J48, via the open source data mining software, WEKA (Waikato Environment for Knowledge Analysis) (Hall, E, Holmes, Pfahringer, Reutemann, & Witten, 2009). J48 is a Java implementation of C4.5 tree algorithm developed by Ross Quinlan (1993) in WEKA.

5.3.3.2 Visualization of trajectory cluster distribution

Visualization techniques integrate human visual pattern acuity and knowledge into the KDD process. They greatly help data mining processes because the human visual system is extremely effective at recognizing patterns, trends, and anomalies (Miller & Han, 2009).

This study employs two visualization techniques to visually confirm the quality of trajectory clustering. The first technique is mapping temporal cluster distributions on a 2D bitmap image. On the map, the x axis represents time, the y axis represents each ID of a mobile object, and each pixel is colored by Cluster ID (Figure 33). This is useful to see if regular and/or irregular patterns of behaviors explained by clustering IDs exist throughout specified time intervals across trajectories of mobile objects. For example, this could be used to explore

similarity and dissimilarity of individual daily activities by looking at distribution patterns of trajectory clustering.

As the second technique, spatio-temporal cluster distributions can be mapped in a 3D space-time cube where the x-y axis represents geographical positions and the z axis to represents time. As described in Chapter 4 in detail, two approaches are used: Space-Time Paths (STPs) and Space-Time Kernel Density Estimate (STKDE). A STP is an individual's trajectory as it resides in a space-time cube, and sub-trajectories of a mobile object can be mapped as a STP with color variations by clustering results. STKDE is a technique to calculate a density distribution in a space-time cube, and cluster distributions of sub-trajectories can be mapped by estimating a line density for each cluster ID using a volume rendering technique. These visualizations are useful to display and explore how human activities regarding to motion behaviors are distributed in space and time.

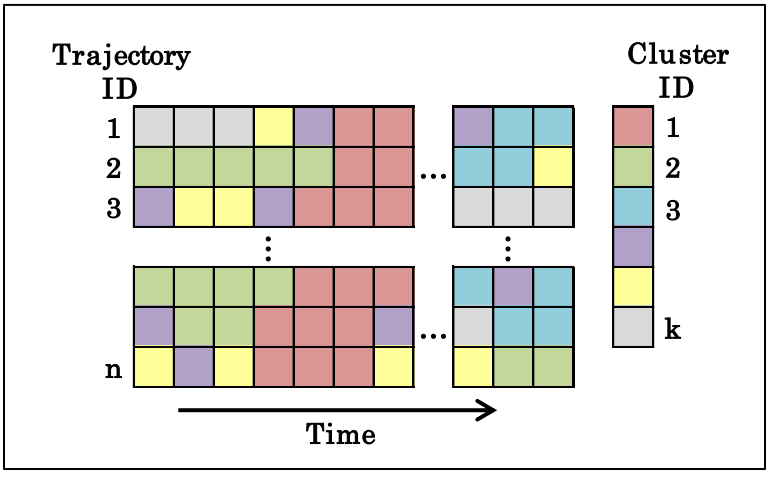

Figure 33. Mapping temporal cluster distribution.

5.3.3.3 Trajectory data mining tool

Figure 34 shows the Graphic User Interface (GUI) of the trajectory data mining tool. The motivation of designing the GUI is to provide the efficiency and ease of use for the underlying trajectory data mining framework developed in this study. Specifically, it includes five components; database and input data selection (blue region), parameter settings for trajectory partitioning (red region) and for trajectory clustering (green region), and output options for results (pink region) and images (yellow region) (Figure 34). The tool enables a user to easily access to a data table containing data of mobile objects stored in the user's MySQL database. A user can also easily select various methodological options (e.g., choice between TRACLUS and Distance-Threshold for trajectory partitioning) and set parameters (e.g., selection of input variables for trajectory clustering). Finally, the tool offers a user to select whether or not to output analytical results as well as images of clustered sub-trajectories.

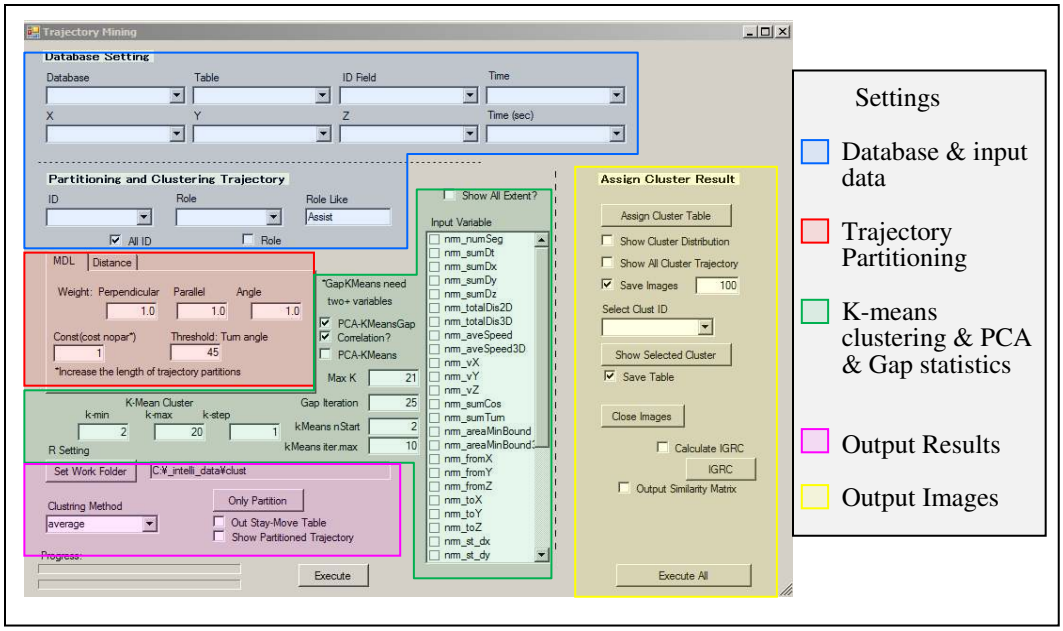

Figure 34. GUI of the trajectory data mining tool (partitioning & clustering).

5.4 Results

To prove how the proposed methodological scheme can be put to use, I now explain how three research challenges can be answered through the results of trajectory data mining with two movement datasets. To recall, three challenges are; 1) how to characterize and generalize massive trajectories to extract interesting patterns; 2) how to explain behavioral contexts of trajectories by those extracted patterns; and 3) how to visualize extracted patterns to overview and compare patterns and trends in space and time.

Two movement datasets were analysed in this study; 1) mixed movements generated by three different random walk models, Brownian Motion, Correlated Random Walk, and Lévy flight; and 2) real-world human movements in urban space collected by a GPS device.

To answer the first research question, the effects of three different trajectory partitioning approaches were examined. Trajectory partitioning is one of the key methodological elements in this work because different partitioning approaches may reveal different behavioral contexts. Here, three partitioning algorithms were compared, no-partitioning, TRACLUS-MDL, and Distance-Threshold. For the second research question, global behavioral contexts of moving objects were reconstructed from extracted clusters of local movement behaviors by the decision tree supervised learning technique. To examine the effect of three partitioning algorithms, behavioral recognition accuracy for each algorithm were compared. To answer the third research question, movement behavioral patterns and process in space and time were examined by visualizing temporal and spatio-temporal trajectory cluster distributions.

5.4.1 Trajectory Data Mining on Simulated Data

5.4.1.1 Dataset

I generated a trajectory dataset with known behaviors to examine the capability of the proposed trajectory data mining framework. It consists of three different movement behaviors generated by three random walk models simulated via R (R Development Core Team, 2008) using the *adehabitat* package (Calenge, 2006). The three models are Brownian Motion (BM), Correlated Random Walk (CRW), and Lévy flight.

BM is considered to be a process of stochastic random walks in a continuous time. In the *adehabitat* package, the process of BM is represented by the function,

$$\frac{1}{h} * B2(t * h^2)$$

where *h* is a scaling parameter for the Brownian motion, *t* is a simulation time step, and $B2_{(t)} = (Bx_{(t)},\ By_{(t)})$ represents a vector of a bivariate Brownian motion, the process of which is normally distributed with mean equal to 0 and variance equal to 1. For BM, *h* is set to 20.

CRW is a random walk where a distribution of turning angle is concentrated. In the model, at each simulation step, the orientation of the move of an agent is drawn from a wrapped normal distribution with concentration parameter *r*, while the length of the move is drawn from a chi distribution multiplied by following,

$$h * \sqrt{dt}$$

where *h* is a scaling parameter (Calenge, 2006). If *r* equals 0, the model generates results similar to BM. For CRW, *r* is set to 0.5 and *h* is set to 20.

Lévy flight is another type of random walk that has a power-law/long-tail distribution of displacement. In the model, at each simulation step, the orientation of move of an agent is drawn from a uniform distribution (−pi, pi), while the length of the move is generated by following,

$$dt * \left(l_0 * \Pr^{\left(\frac{1}{(1-mu)}\right)} \right)$$

where l_0 is the minimum length of a step, Pr is a uniform distribution function drawing a random value between 0 and 1, *mu* is the exponent of the Lévy distribution. For Lévy flight, l_0 is set to 10 and *mu* is set to 2.2.

For each model, the number of agents is set to 100 and each simulation was run for 400 simulation steps. The results of trajectories from BM, CRW, and Lévy flight are shown in Figure 35, Figure 36, and Figure 37 respectively. In addition, I merged the three datasets into one (n=300) (Figure 38), and the trajectory data mining tool was executed with the mixed trajectory dataset, which consists of three different movement behaviors (BM, CRW, and Lévy flight).

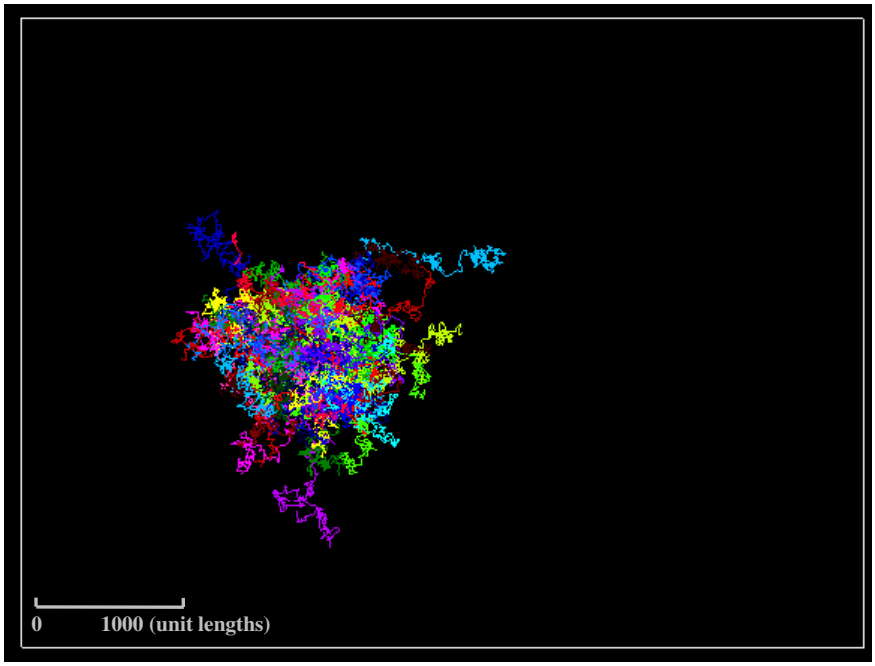

Figure 35. Trajectories of BM (n=100, t = 400, colored randomly by trajectory ID).

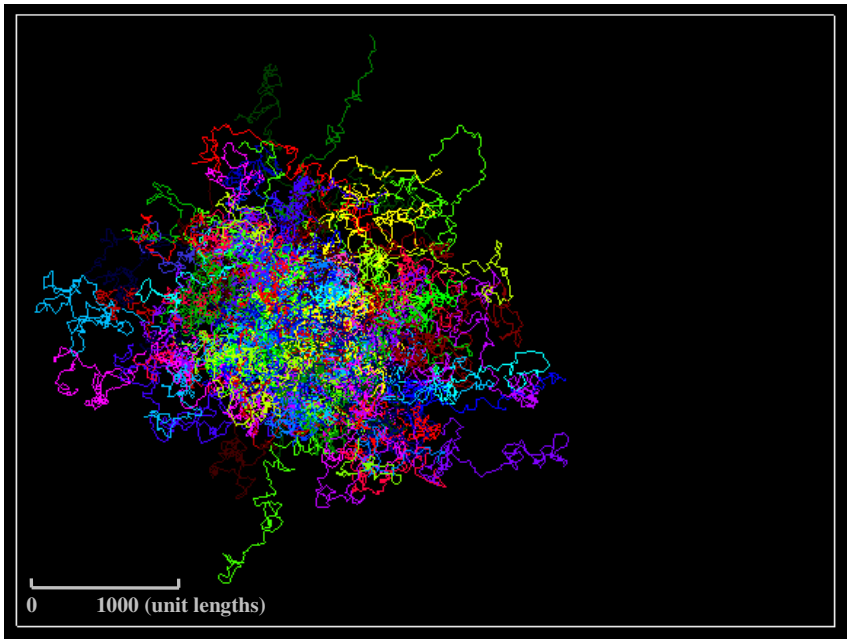

Figure 36. Trajectories of CRW (n=100, t = 400, colored randomly by trajectory ID).

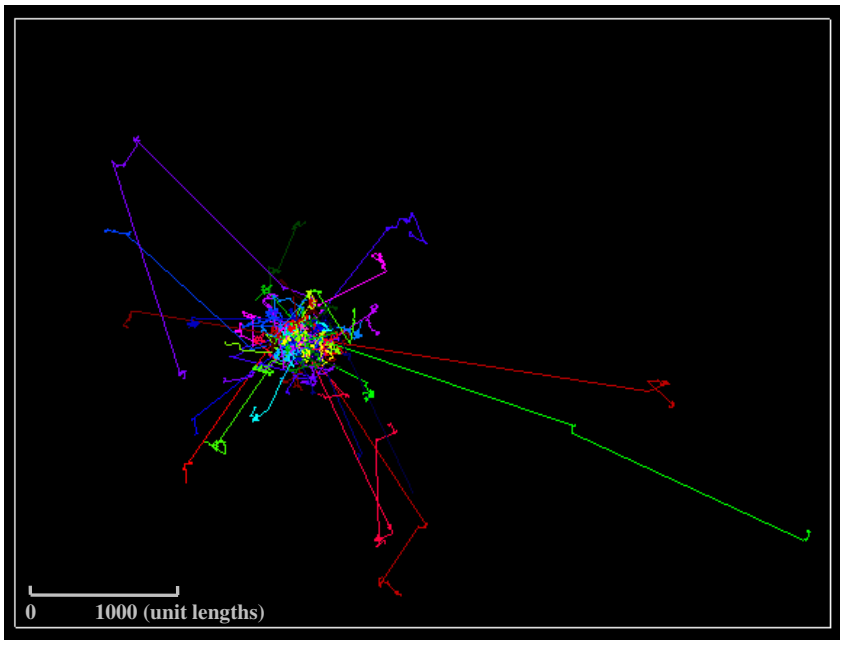

Figure 37. Trajectories of Lévy flight (n=100, t = 400, colored randomly by trajectory ID).

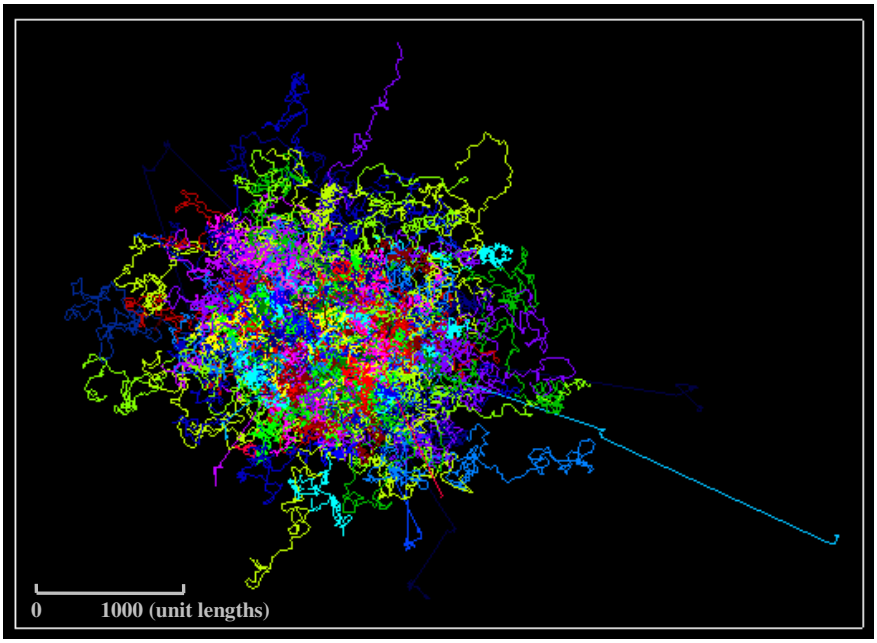

Figure 38. Mixed trajectories of BM, CRW, and Lévy flight (n= 300, t = 400).

5.4.1.2 Results

To answer the first research question of how to characterize and generalize massive trajectories to extract interesting patterns, I compared and examined three different trajectory partitioning algorithms; no-partitioning, TRACLUS-MDL, and Distance-Threshold.

Figure 39, Figure 40, and Figure 41 represent the results of trajectory partitioning for three random walk simulations by two partitioning algorithms. The parameter value of c was set to 0.75 in the TRACLUS-MDL approach, while Th_d, was set to 20 (unit lengths) in the Distance-Threshold approach. In the following figures, each partitioned trajectory in a trajectory is alternately colored by red and cyan. Table 5 shows the number of sub-trajectories in each partition algorithm and the percentage of data compression calculated as (1 – number of

sub-trajectories / total number of segments) * 100, where the total number of segments equals to 40,000 (= 100 agents times 400 simulation time steps) for each simulation. In BM and CRW, Distance-Threshold has a slightly large number of sub-trajectories than TRACLUS-MDL; however, as shown in Figure 39 and Figure 40, partitioning patterns are very different.

To recall the difference of two approaches, the TRACLUS-MDL approach partitions a trajectory by finding a sudden geometrical change, which essentially takes the directionality of movements into account. On the contrary, the Distance-Threshold approach partitions a trajectory by finding a slow movement, labeled as *STAY*, determined such that a sub-trajectory distance is less than a defined value of distance threshold. Because each simulation time step is the same, the Distance-Threshold approach identifies a segment with slow movement or staying behavior and partitions a trajectory at the segment. One can see several key differences between the two partitioning approaches in terms of movement behaviors in each random walk model. In BM and Lévy flight, when using TRACLUS-MDL, trajectories were frequently partitioned because the orientation of move of an agent was randomly drawn (Figure 39 and Figure 41). On the other hand, trajectories of CRW were less frequently partitioned by TRACLUS-MDL because the orientation is drawn from a wrapped normal distribution that concentrated the turning angle of an agent (Figure 40). TRACLUS-MDL is useful to partition trajectories like CRW if they are composed of some directed movement and if their behavior changes with the change in movement directionality.

While TRACLUS-MDL considers directionality, Distance-Threshold takes the length of sub-trajectories into account. In BM and CRW, the short length of segment, which is less than the distance threshold, was caused by the probability based on bivariate normal distribution (BM) and Chi distribution (CRW). Because of this, a long directed sub-trajectory in CRW can be partitioned at the middle of the path (Figure 40). To the contrary, Lévy flight provided many short movements because the step size of an agent followed a power-law distribution, and thus Distance-Threshold partitioning separated a cluster of small movements and very long steps. This suggests that if movement behaviors of a mobile object are composed of stay and move behaviors, the Distance-Threshold approach can distinguish the two different movements (Figure 42).

Table 5. Number of sub-trajectories in each partitioning algorithm (Simulation).

	BM		CRW		Lévy Flight	
	sub-TRs	compress	sub-TRs	compress	sub-TRs	compress
TRACLUS-MDL	18232	54.42%	16257	59.36%	11652	70.87%
Distance-Threshold	19045	52.39%	18853	52.87%	2137	94.66%

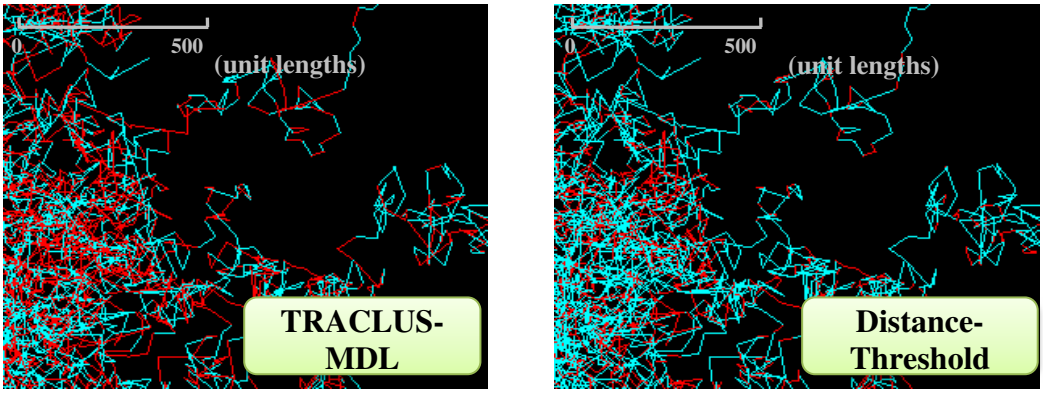

Figure 39. Trajectory partitioning results of Brownian motion.

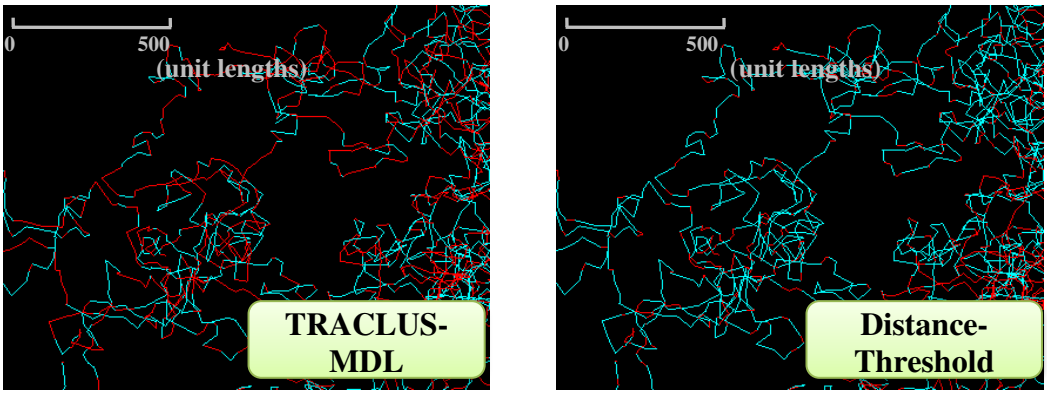

Figure 40. Trajectory partitioning results of Correlated Random Walk.

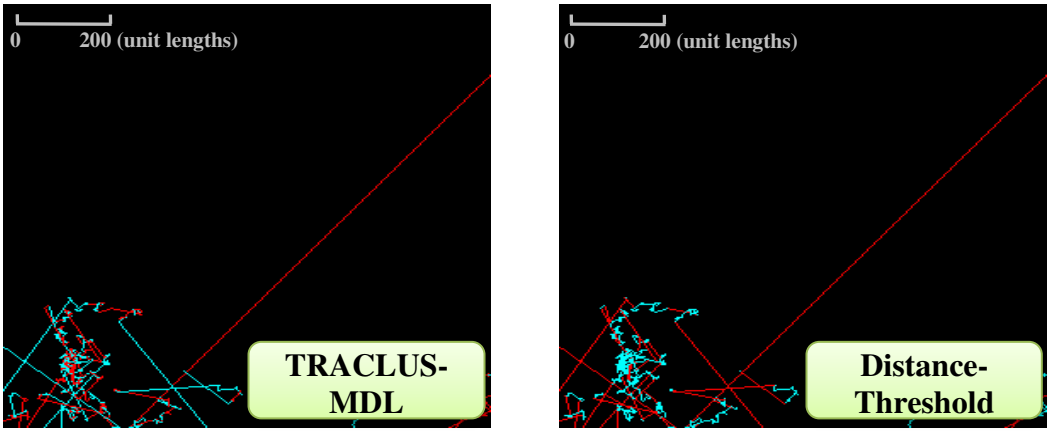

Figure 41. Trajectory partitioning results of Lévy Flight.

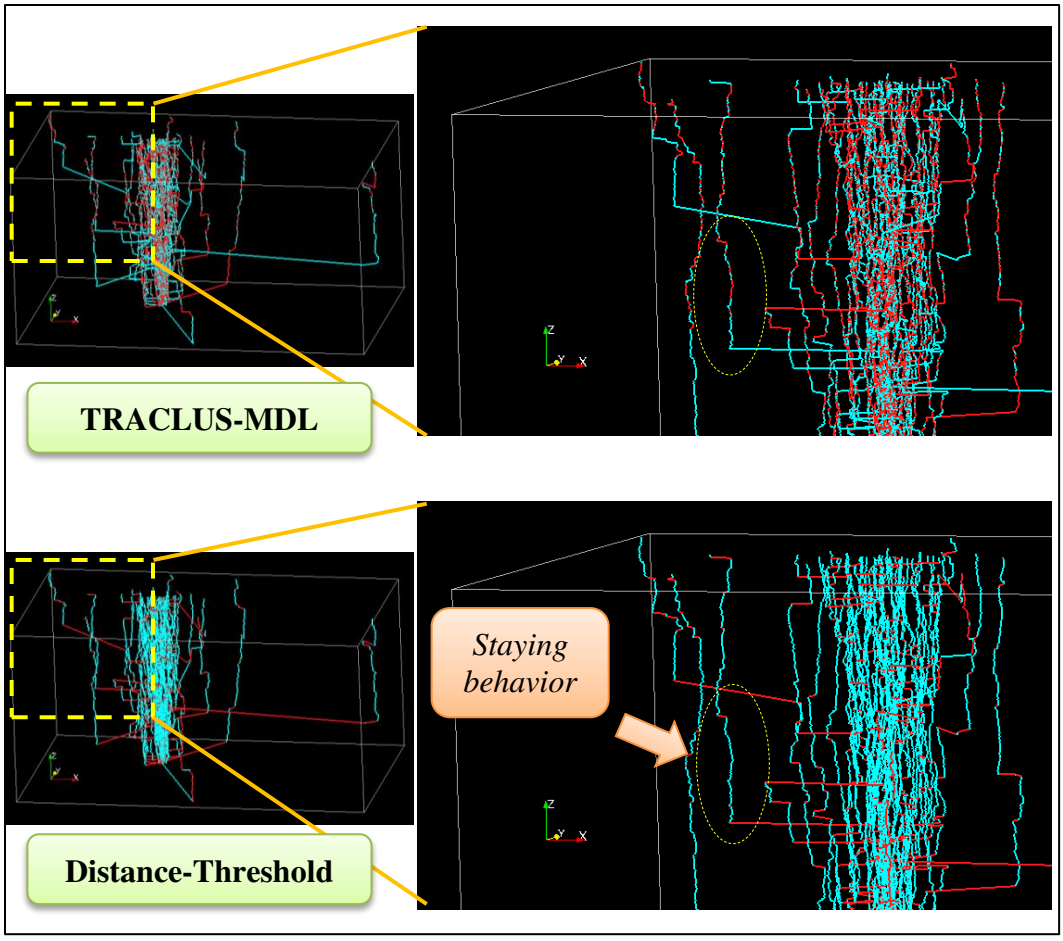

Figure 42. Trajectory partitioning results of Lévy Flight in Space-Time Cube. Distance-Threshold partitioning can capture a long staying behavior that is composed of many short movements.

For each trajectory partition in the three different partition algorithms, multi-dimensional vectors were calculated to characterize the partition trajectory. The multiple metrics can describe complex movement behaviors, which cannot be explained by just a single variable. To identify dependencies of multiple motion variables in each sub-trajectory dataset, correlation analysis was performed. Table 6 to Table 8 detail correlation matrices for movement variables of trajectory

partition in the three partition algorithms respectively. In the no-partition algorithm, where trajectories are not partitioned, each trajectory has the same duration (dt=400); therefore, dx and vx, dy and vy, and $d2D$ and v are perfectly correlated. In addition, only positive correlations have been observed in the no-partition algorithm because it treats a trajectory as a whole and each trajectory describes the diffusing process of random walks. Between TRACLUS-MDL and Distance-Threshold, large differences are found in the correlation between duration (d_t) and distance variables (d_x, d_x, $d2D$, d_{sx}, d_{sx}, d_s2D) and the correlation between velocity variables (v_x, v_x, v) and minimum boundary box (mbb). While TRACLUS-MDL has positive correlations between duration and distance variables between start and end nodes, Distance-Threshold has no correlation. Sub-trajectories by Distance-Threshold contain both moving and staying behavior. While the property of moving behavior has a positive correlation between duration and travel length, the property of staying behavior shows a large duration with short travel length; therefore, no correlation between duration and distance variables is identified in the Distance-Threshold partitioning. On the other hand, sub-trajectories by TRACLUS-MDL can be composed of moving behavior, staying behavior, or both behaviors because TRACLUS-MDL considers directionality rather than distance. Therefore, those sub-trajectories show no correlation between velocity variable and minimum boundary box.

This finding implies that TRACLUS-MDL is useful when a research objective is to find mobile objects' behavior due to their directional change (e.g., identifying normal/abnormal patterns of hurricane trajectories for track prediction

and hazards prevention; finding seasonal migration patterns of animals for geo-behavioral studies and conservation purposes). On the contrary, because partitioned trajectories by TRACLUS-MDL may contain mixed patterns of staying and moving activities, it will be difficult to distinguish such behaviors (e.g., human daily behavior involving various activities of staying at, for example, a home, an office, and stores, and of moving such as commuting and shopping), which the Distance-Threshold approach specifically focuses on extracting such behaviors.

As shown in the correlation matrix, some variables can be highly correlated in sub-trajectory datasets and such variables provide only redundant information when performing clustering analysis. To reduce the dimensionality of the dataset consisting of interrelated variables, PCA was conducted.

Table 9, Table 10, and Table 11 represent the results of PCA for three partitioning algorithms. The numbers of identified Principal Components (PCs) with eigenvalue greater than 1 are 2, 4, and 3 for no-partition, TRACLUS-MDL, and Distance-Threshold respectively. These PCs explain 86.1%, 89.5%, and 82.1% of the original variables for the dataset in the three partitioning approaches respectively.

Table 6. Correlation matrix for movement variables (Sim: no partition).

	dx	*dy*	*d2D*	*v*	*vx*	*vy*	*dsx*	*dsy*	*ds2D*	*mbb*	*sct*
dx	1										
dy	0.9873	1									
d2D	0.9969	0.9965	1								
v	0.9969	0.9965	1	1							
vx	1	0.9873	0.9969	0.9969	1						
vy	0.9873	1	0.9965	0.9965	0.9873	1					
dsx	0.2627	0.2152	0.2393	0.2393	0.2627	0.2152	1				
dsy	0.2635	0.268	0.2640	0.2640	0.2635	0.2680	0.3097	1			
ds2D	0.3283	0.2958	0.3116	0.3116	0.3283	0.2958	0.8679	0.7233	1		
mbb	0.4206	0.3956	0.4075	0.4075	0.4206	0.3956	0.7525	0.6305	0.8523	1	
sct	0.5047	0.4993	0.5033	0.5033	0.5047	0.4993	0.3311	0.3089	0.4023	0.5345	1

Table 7. Correlation matrix for movement variables (Sim: TRACLUS-MDL).

	dt	*dx*	*dy*	*d2D*	*v*	*vx*	*vy*	*dsx*	*dsy*	*ds2D*	*mbb*	*sct*
dt	1											
dx	0.6767	1										
dy	0.6915	0.6998	1									
d2D	0.7429	0.9218	0.9163	1								
v	-0.0384	0.3912	0.4008	0.4323	1							
vx	-0.0310	0.4805	0.1480	0.3379	0.7813	1						
vy	-0.0298	0.1443	0.4911	0.3365	0.7819	0.255	1					
dsx	0.4967	0.8575	0.5060	0.7399	0.3781	0.4892	0.1103	1				
dsy	0.5137	0.5170	0.8371	0.7283	0.3972	0.1183	0.5107	0.4198	1			
ds2D	0.6020	0.8273	0.7928	0.8793	0.4664	0.3685	0.3602	0.8495	0.8211	1		
mbb	0.3894	0.5452	0.4498	0.5278	0.0673	0.0616	0.0492	0.5947	0.4761	0.6259	1	
sct	-0.0283	0.0724	0.0788	0.0822	0.0480	0.0392	0.0355	0.2420	0.2753	0.3101	0.0071	1

Table 8. Correlation matrix for movement variables (Sim: Distance-Threshold).

	dt	*dx*	*dy*	*d2D*	*v*	*vx*	*vy*	*dsx*	*dsy*	*ds2D*	*mbb*	*sct*
dt	1											
dx	0.3521	1										
dy	0.3712	0.7621	1									
d2D	0.3868	0.9422	0.9299	1								
v	-0.0616	0.5822	0.5032	0.5770	1							
vx	-0.0510	0.6405	0.3325	0.5207	0.9136	1						
vy	-0.0577	0.3683	0.5990	0.4980	0.8540	0.5837	1					
dsx	0.0549	0.7946	0.5065	0.6983	0.7322	0.8046	0.4614	1				
dsy	0.0592	0.5269	0.7703	0.6804	0.6494	0.4344	0.7664	0.4778	1			
ds2D	0.0668	0.7822	0.7327	0.8083	0.8084	0.7350	0.6923	0.8731	0.8325	1		
mbb	0.0188	0.5446	0.4740	0.5321	0.8571	0.8023	0.7446	0.6666	0.5923	0.7205	1	
sct	-0.0793	0.0577	0.0622	0.0638	0.0239	0.0207	0.0218	0.1943	0.2052	0.2345	0.0560	1

Table 9. Results of PCA (Sim: No Partition).

Variables	Principal Components		
	Loadings		Contribution
	1	2	
d_x	-0.9609	-0.2587	0.9902
d_y	-0.9527	-0.2896	0.9914
d_{2D}	-0.9593	-0.2767	0.9968
v	-0.9593	-0.2767	0.9968
v_x	-0.9609	-0.2587	0.9902
v_y	-0.9527	-0.2896	0.9914
d_{sx}	-0.4459	0.7223	0.7206
d_{sy}	-0.4426	0.5830	0.5358
d_{s2D}	-0.5493	0.8087	0.9557
mbb	-0.6285	0.6930	0.8753
sct	-0.6232	0.1923	0.4254
Eigen.values	6.9820	2.4877	9.4696
Proportion	63.47	22.62	86.09
Cumulative.prop.	63.47	86.09	-

Table 10. Results of PCA (Sim: TRACLUS-MDL).

Variables	Principal Components				
	Loadings				Contribution
	1	2	3	4	
d_t	-0.6311	-0.5645	-0.0870	-0.1926	0.7617
d_x	-0.9015	-0.1336	0.3084	-0.1108	0.9380
d_y	-0.8838	-0.0738	-0.3763	-0.0389	0.9297
d_{2D}	-0.9636	-0.1132	-0.0214	-0.0825	0.9486
v	-0.5562	0.8134	0.0130	-0.1221	0.9860
v_x	-0.4611	0.6608	0.4883	-0.1642	0.9147
v_y	-0.4610	0.6896	-0.4669	-0.0415	0.9078
d_{sx}	-0.8407	-0.1056	0.4472	0.0626	0.9218
d_{sy}	-0.8220	-0.0237	-0.4379	0.1836	0.9018
d_{s2D}	-0.9675	-0.0758	0.0384	0.1383	0.9624
mbb	-0.6645	-0.3962	0.0643	0.0267	0.6034
sct	-0.1498	0.1391	0.1141	0.9539	0.9647
Eigen.values	6.4378	2.1224	1.1117	1.0686	10.7406
Proportion	53.65	17.69	9.26	8.91	89.51
Cumulative.prop.	53.65	71.34	80.60	89.51	-

Table 11. Results of PCA (Sim: Distance-Threshold).

Variables	Principal Components			
	Loadings			Contribution
	1	2	3	
d_t	0.1552	0.7901	-0.2076	0.6914
d_x	0.8354	0.3759	-0.0802	0.8457
d_y	0.7917	0.4697	0.0853	0.8547
d_{2D}	0.8632	0.4546	0.0005	0.9518
v	0.8856	-0.3802	-0.1578	0.9538
v_x	0.8042	-0.3711	-0.2347	0.8396
v_y	0.7741	-0.3045	-0.0262	0.6927
d_{sx}	0.8418	-0.0731	0.0136	0.7141
d_{sy}	0.8017	0.0292	0.2706	0.7168
d_{s2D}	0.9553	-0.0227	0.1583	0.9381
mbb	0.8228	-0.3196	-0.1453	0.8003
sct	0.1304	-0.0554	0.9200	0.8664
Eigen.values	7.0823	1.6797	1.1034	9.8654
Proportion	59.02	14.00	9.19	82.21
Cumulative.prop.	59.02	73.02	82.21	-

PC scores of each sub-trajectory for each PC (Eigen value ≥ 1) were calculated, and then used as a new input dataset for cluster analysis. K-means clustering was run for each sub-trajectory dataset in three partition algorithms with different *k* in a range between 2 and 20, which is arbitrarily defined. The optimal values of *k* were estimated by applying the gap statistic, by identifying $\hat{k} = smallest\ k\ such\ that\ Gap(k) \geq Gap(k+1) - s(k+1)$. The number of generating reference datasets of a null model, *B*, was set to 25. Figure 43 illustrates gap curves for three partition algorithms, where large dots indicate that *Gap(k)* is greater than or equal to *Gap(k+1) – s(k+1)*. This study also considers the number of *k* determined by the highest value of *Gap(k)* in the range of *k* between 2 and 20 as an alternative value because the highest gap value represents

the largest difference of the compactness of clusters between a raw dataset and a null reference dataset (i.e., random distribution in this study). Following results of cluster analysis and the gap statistics, optimal values of *k* are 3, 2, and 5 for no-partition, TRACLUS-MDL, and Distance-Threshold respectively. *k* values determined by the highest value of *Gap(k)* are 13, 4, and 19 respectively. Figure 44 to Figure 49 show the numbers of sub-trajectories assigned to a cluster for corresponding partitioning methods and selected *k* values.

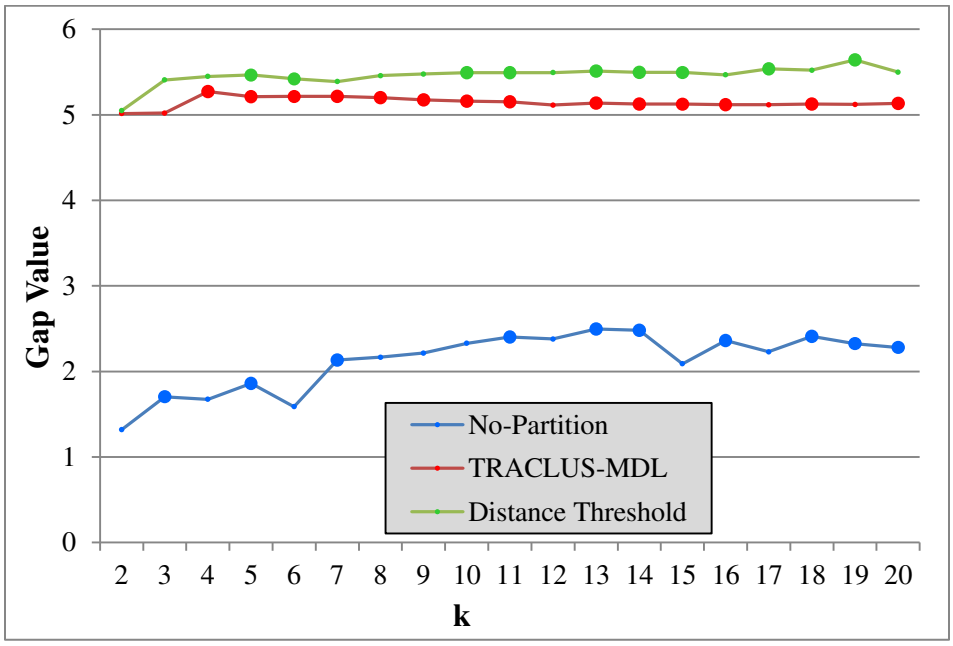

Figure 43. Gap curve for three partitioning algorithms (Simulation).

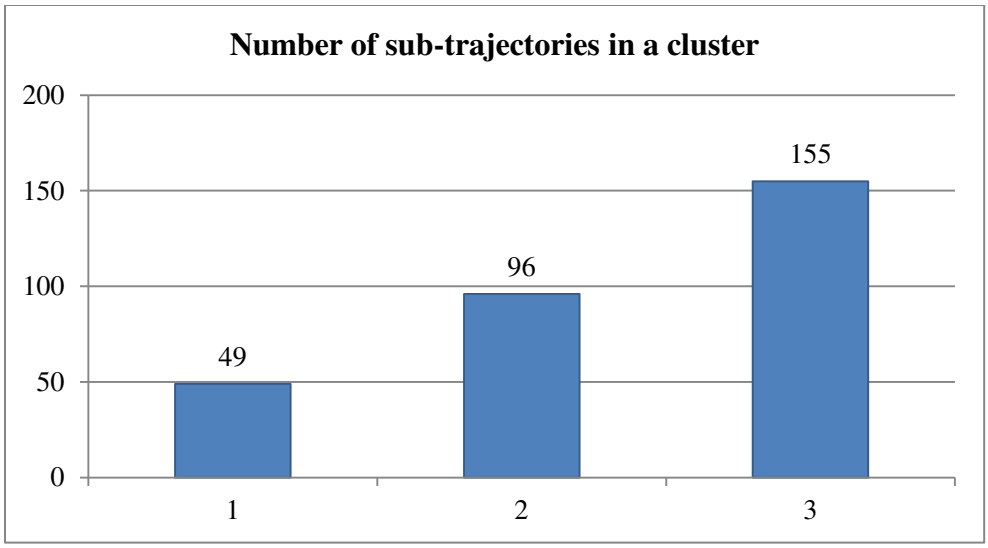

Figure 44. Number of subtrajectories in a cluster (n=300) (no partition: *k*=3).

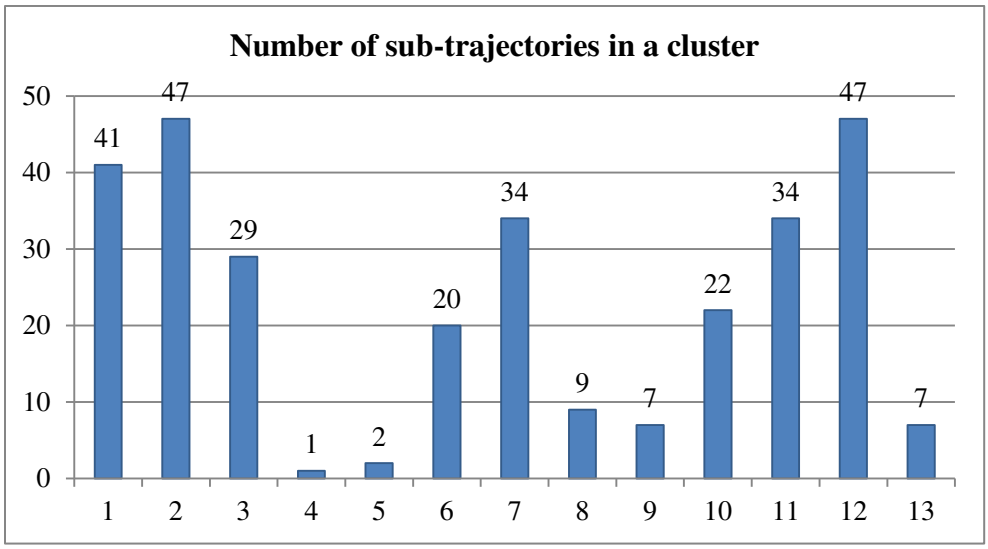

Figure 45. Number of subtrajectories in a cluster (n=300) (no partition: *k*=13).

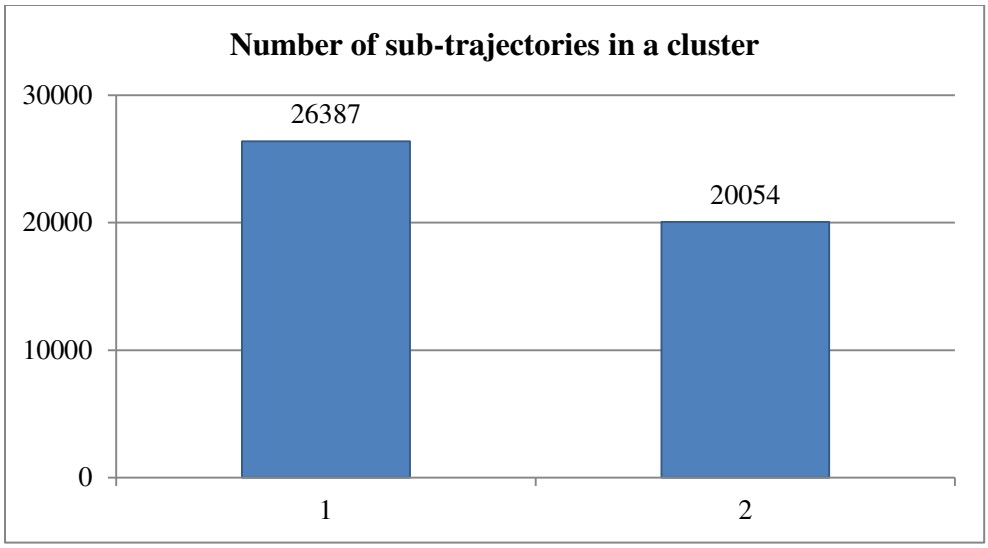

Figure 46. Number of subtrajectories in a cluster (n =46,441) (TRACLUS-MDL: *k*=2).

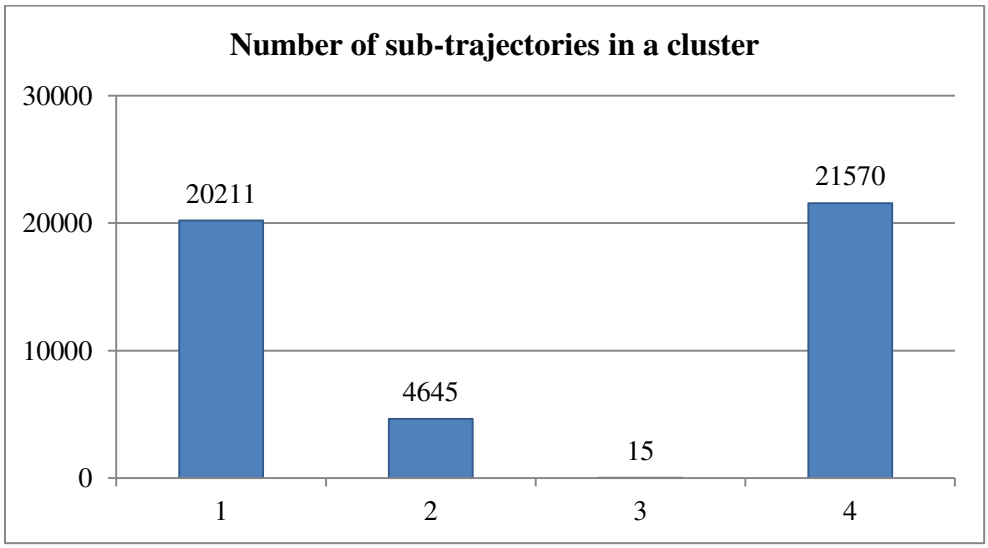

Figure 47. Number of subtrajectories in a cluster (n =46,441) (TRACLUS-MDL: *k*=4).

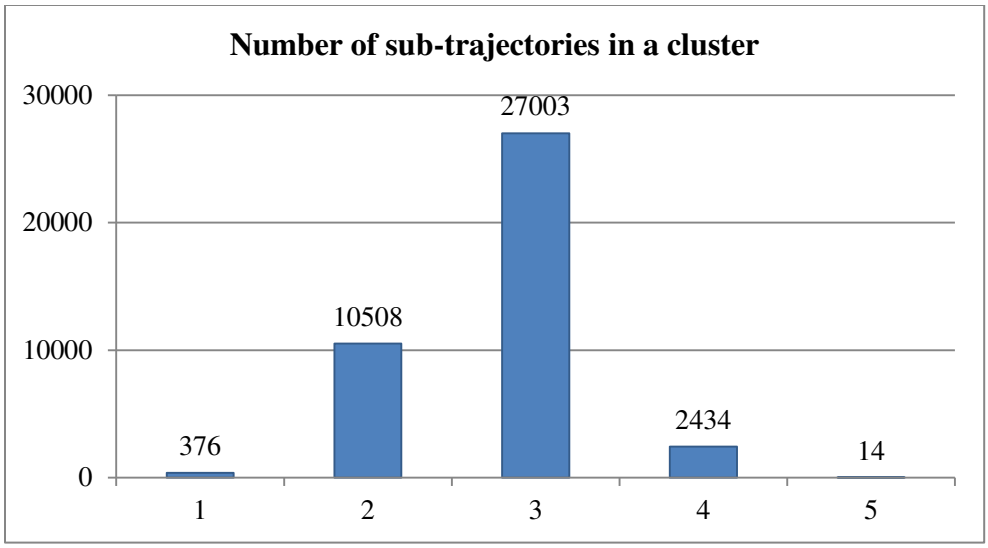

Figure 48. Number of subtrajectories in a cluster (n =40,335) (Distance-Threshold: *k*=5).

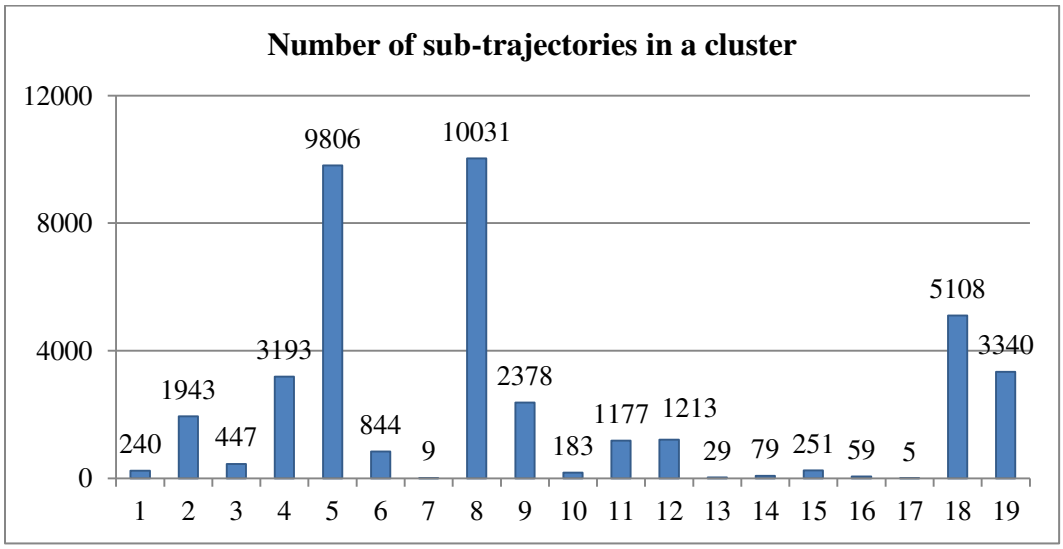

Figure 49. Number of subtrajectories in a cluster (n =40,335) (Distance-Threshold: *k*=19).

Figure 50 to Figure 55 present cluster profiles for corresponding partitioning methods and selected *k* values, where the vertical axis is cluster ID and the horizontal axis shows the average of normalized value of independent variables within a cluster. Figure 56 to Figure 61 display sub-trajectories for each cluster ID for corresponding partitioning methods and selected *k* values. These figures explain sub-trajectory characteristics within a cluster. In the no partition algorithm that treats a trajectory as a whole, the optimal *k* value estimated by the gap statistic was 3. The cluster profile (Figure 50) and the image of trajectories (Figure 56) illustrate that trajectories of Cluster 1 represent long travel distance and directed movement. Trajectories of Cluster 2 are described as short travel distance, slow movement, and sinuous path, where as those of Cluster 3 are described as long travel distance and but sinuosity of those paths is between Cluster 1 and 2. These clusters roughly explain three random walk behaviors, where Cluster 1 is CRW, Cluster 2 is Lévy Flight, and Cluster 3 is BM. Using the highest gap value, 13 trajectory clusters were identified. These clusters classified 3 random walk behaviors into groups in further detail, and some trajectory clusters in Figure 57 explain those behaviors well (e.g., Cluster 4, 5, 8, and 9 for Lévy Flight).

As opposed to the no partitioning approach, trajectory clustering in TRACLUS-MDL and Distance-Threshold approaches classified partitioned sub-trajectories into groups that explain some portion of movement behavior in a trajectory. In TRACLUS-MDL, the optimal *k* value is 2 determined by the gap statistic. In two clusters, one cluster describes longer and less sinuous sub-

trajectories, and the other one is vice-versa (Figure 52 and Figure 58). By the highest gap value, four clusters were identified (Figure 53 and Figure 59). In these clusters, sub-trajectories of Cluster 3 were long directed paths that explain parts of Lévy Flight trajectory, whereas the other three clusters represented short and sinuous sub-trajectories in different degrees of length and sinuosity.

In Distance-Threshold, the optimal k value estimated by the gap statistic was 5 and estimated by the highest value of gap statistic was 19 (Figure 54, Figure 55, Figure 60, and Figure 61). Some of these clusters describe a long trip of Lévy Flight trajectory well (Cluster 5 with $k = 5$, Cluster 7, 13, and 17 with $k = 19$). In addition, some other clusters can describe staying behaviors in different degrees of duration (Cluster 1, 10, 14, 15, and 16), while others explain moving behaviors in different degrees of length, velocity, and sinuosity.

The key difference of partitioned sub-trajectories between TRACLUS-MDL and Distance-Threshold is the treatment of staying behavior. For example, sub-trajectories of Cluster 3 in TRACLUS-MDL with $k = 4$ and Cluster 5 in Distance-Threshold with $k = 5$ have similar shapes and both describe a long trip of Lévy Flight trajectory; however, the key difference is duration. Because Distance-Threshold can differentiate between STAY and MOVE, the sub-trajectory of Cluster 5 only represents MOVE so that its duration is small (Figure 54 and Figure 60). On the contrary, because TRACLUS-MDL does not consider staying behaviors, the sub-trajectory of Cluster 3 by TRACLUS-MDL contains staying behavior with the long trip of Lévy Flight so that its duration is large (Figure 53 and Figure 59).

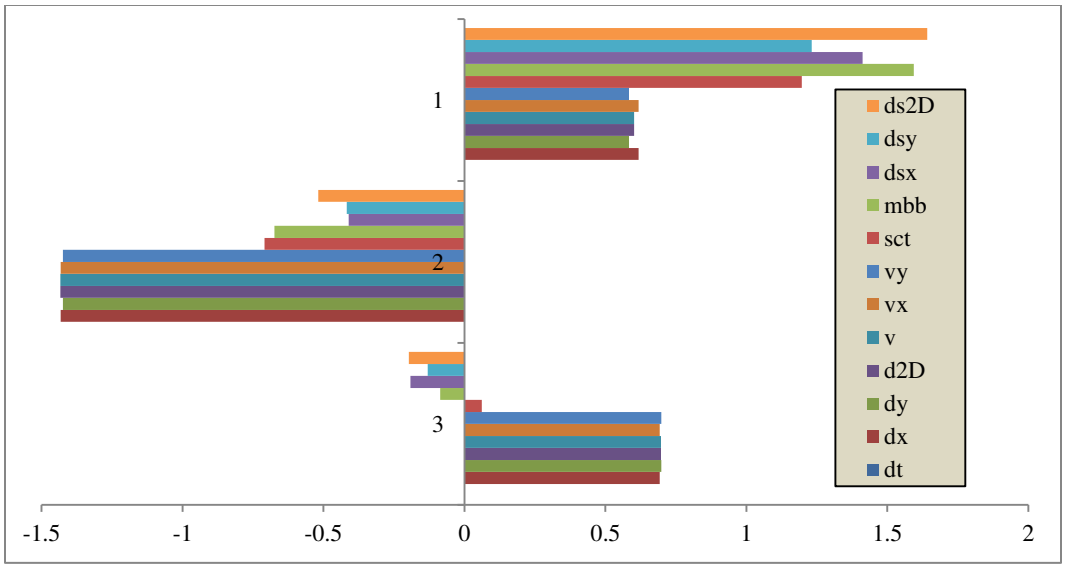

Figure 50. Cluster profile (no-partition: *k*=3).

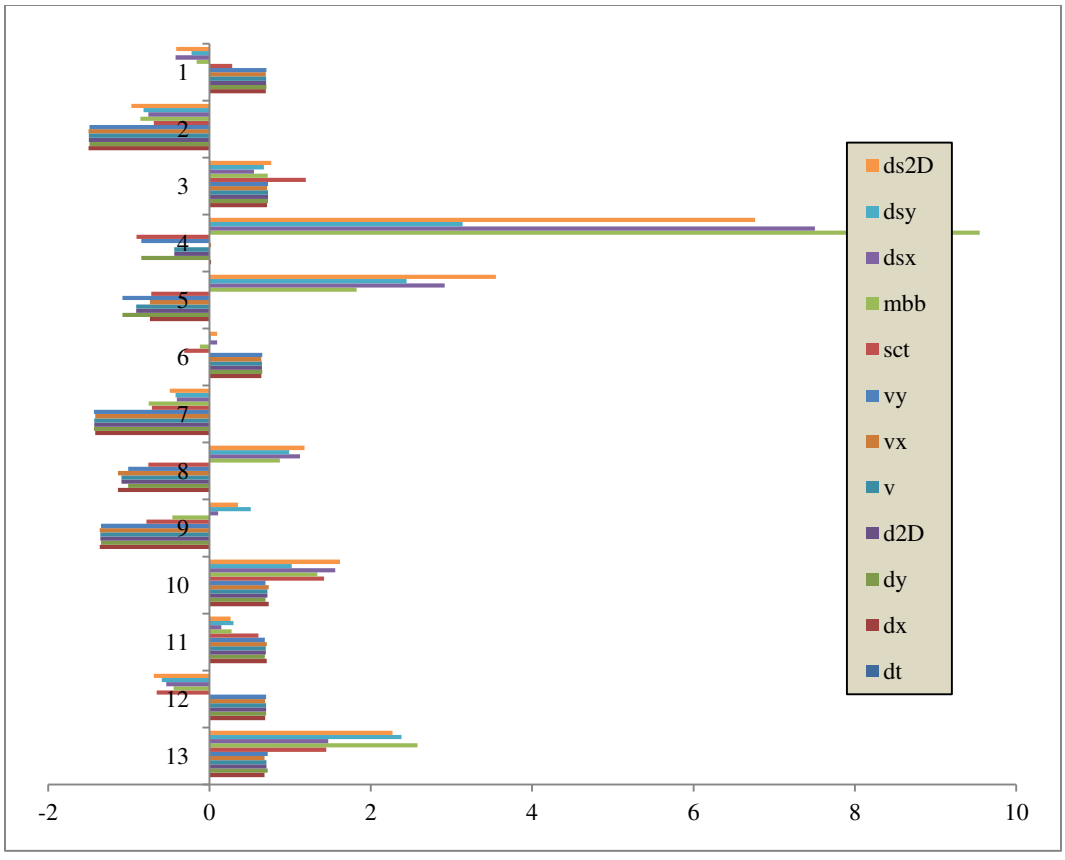

Figure 51. Cluster profile (no-partition: *k*=13).

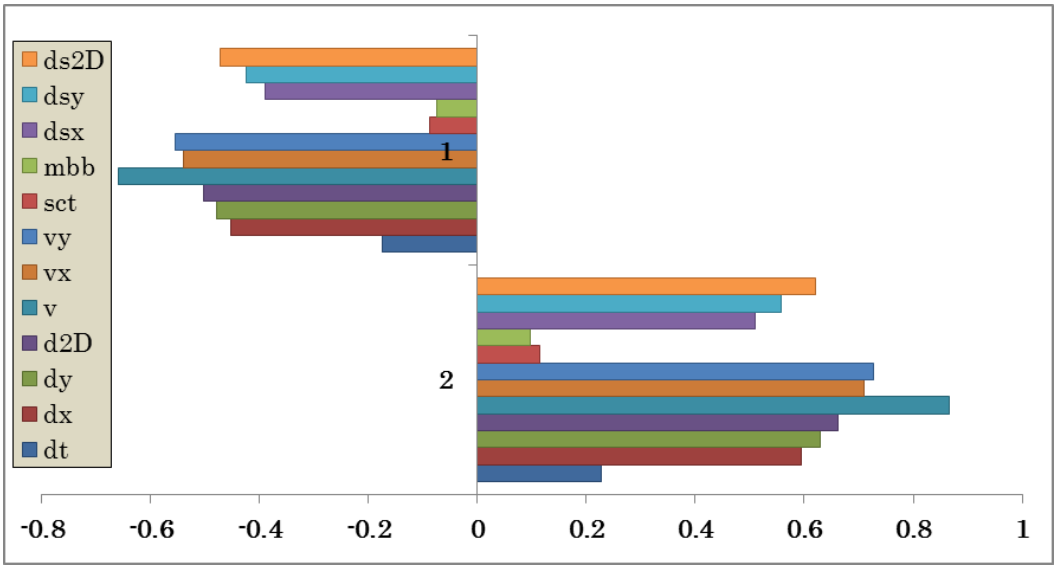

Figure 52. Cluster profile (TRACLUS-MDL: *k*=2).

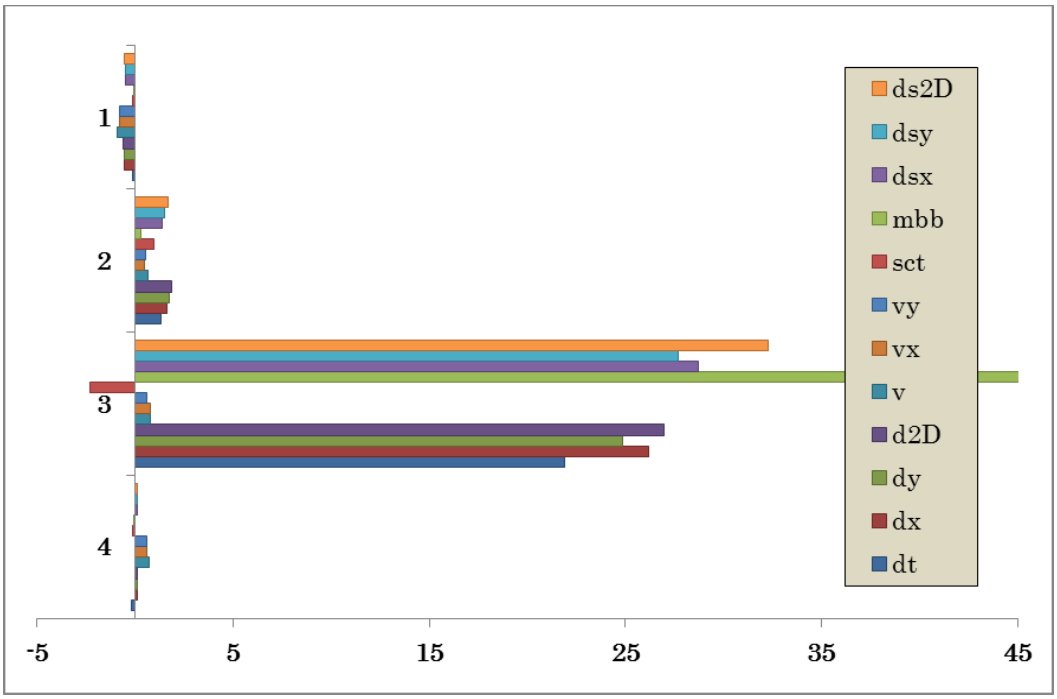

Figure 53. Cluster profile (TRACLUS-MDL: *k*=4).

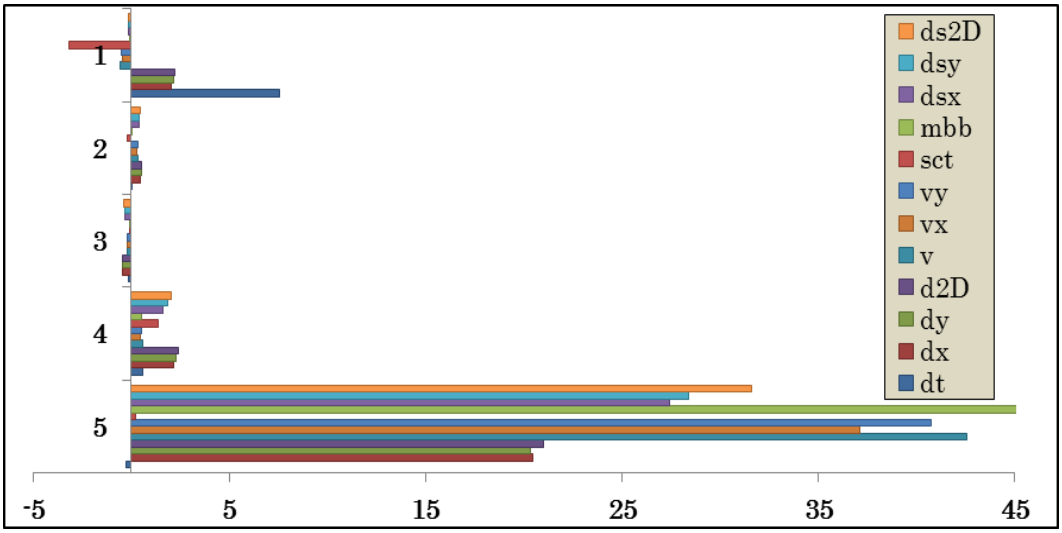

Figure 54. Cluster profile (Distance-Threshold: *k*=5).

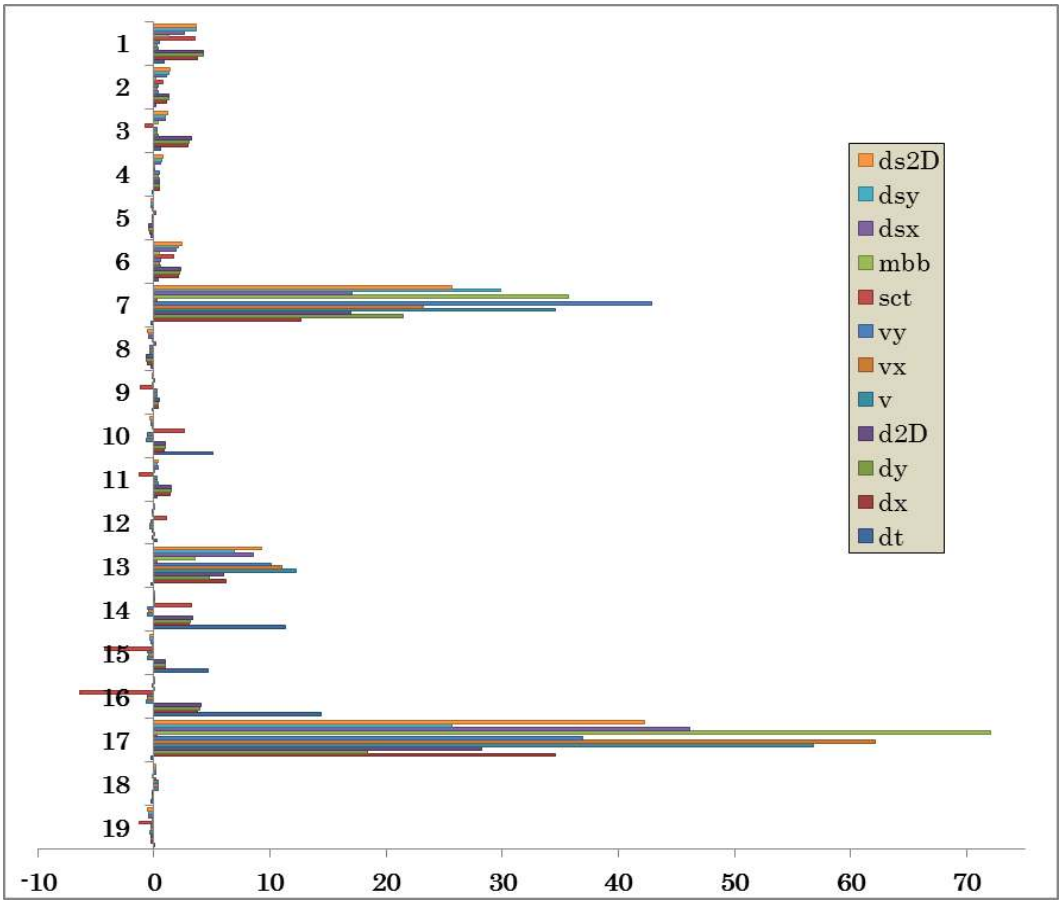

Figure 55. Cluster profile (Distance-Threshold: *k*=19).

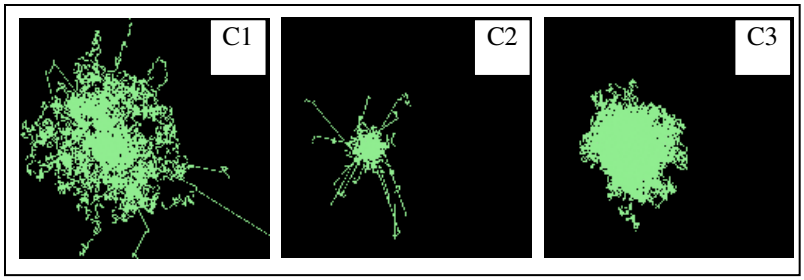

Figure 56. Trajectory clusters (no partition: *k*=3).

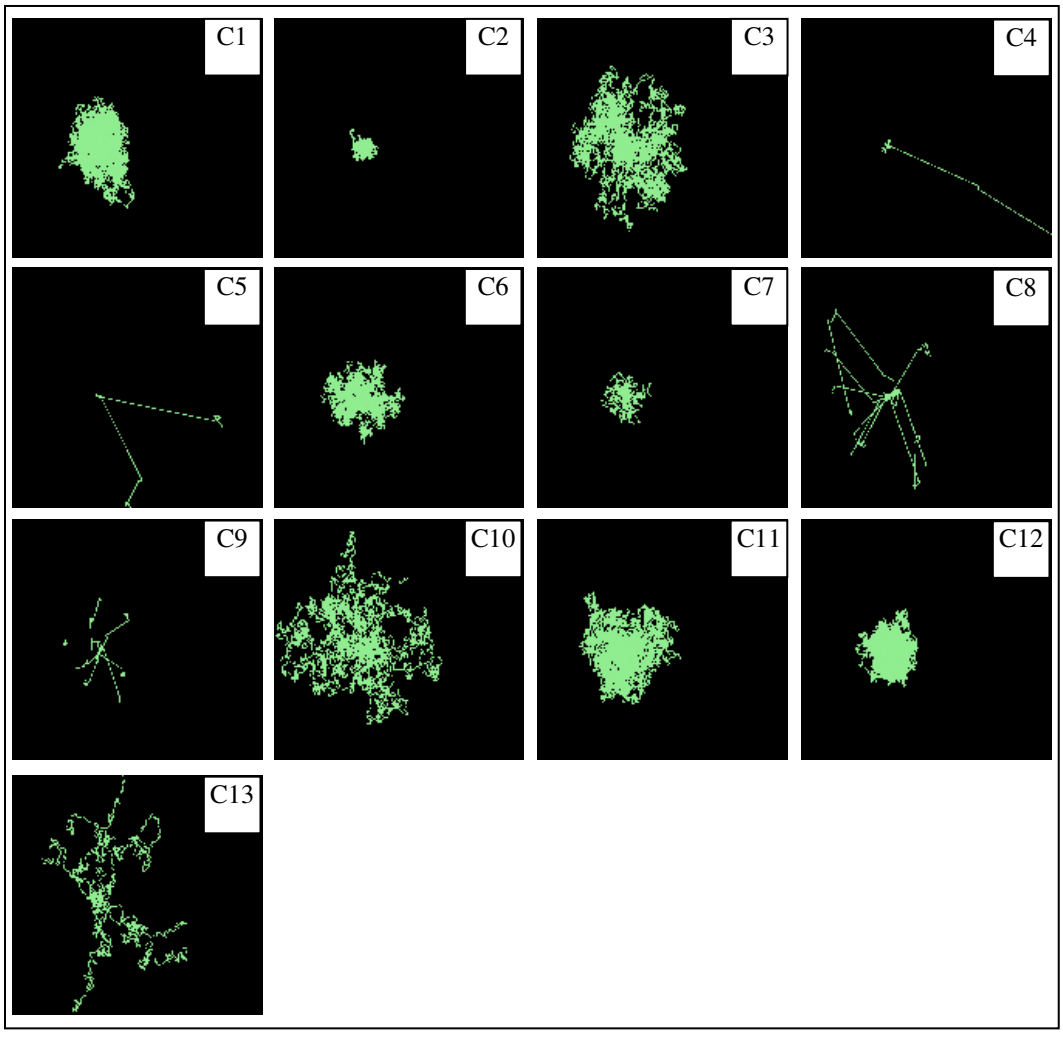

Figure 57. Trajectory clusters (no partition: *k*=13).

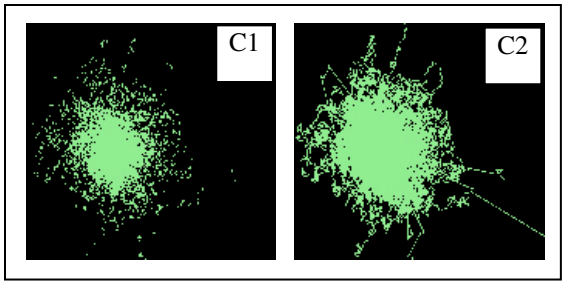

Figure 58. Sub-trajectory clusters (TRACLUS-MDL: *k*=2).

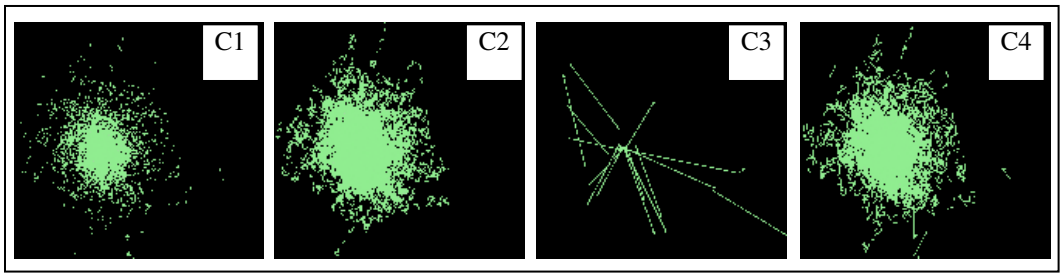

Figure 59. Sub-trajectory clusters (TRACLUS-MDL: *k*=4).

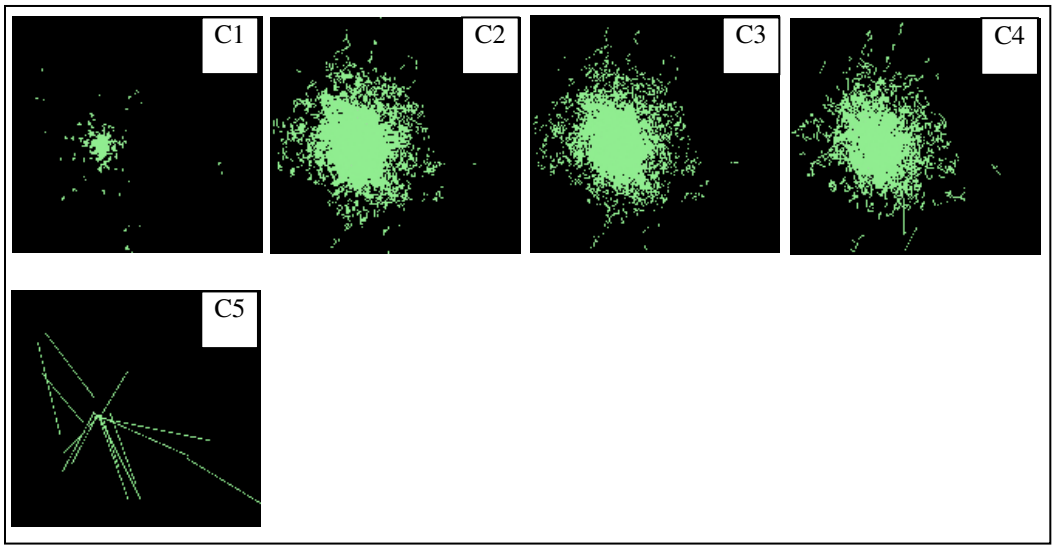

Figure 60. Sub-trajectory clusters (Distance-Threshold: *k*=5).

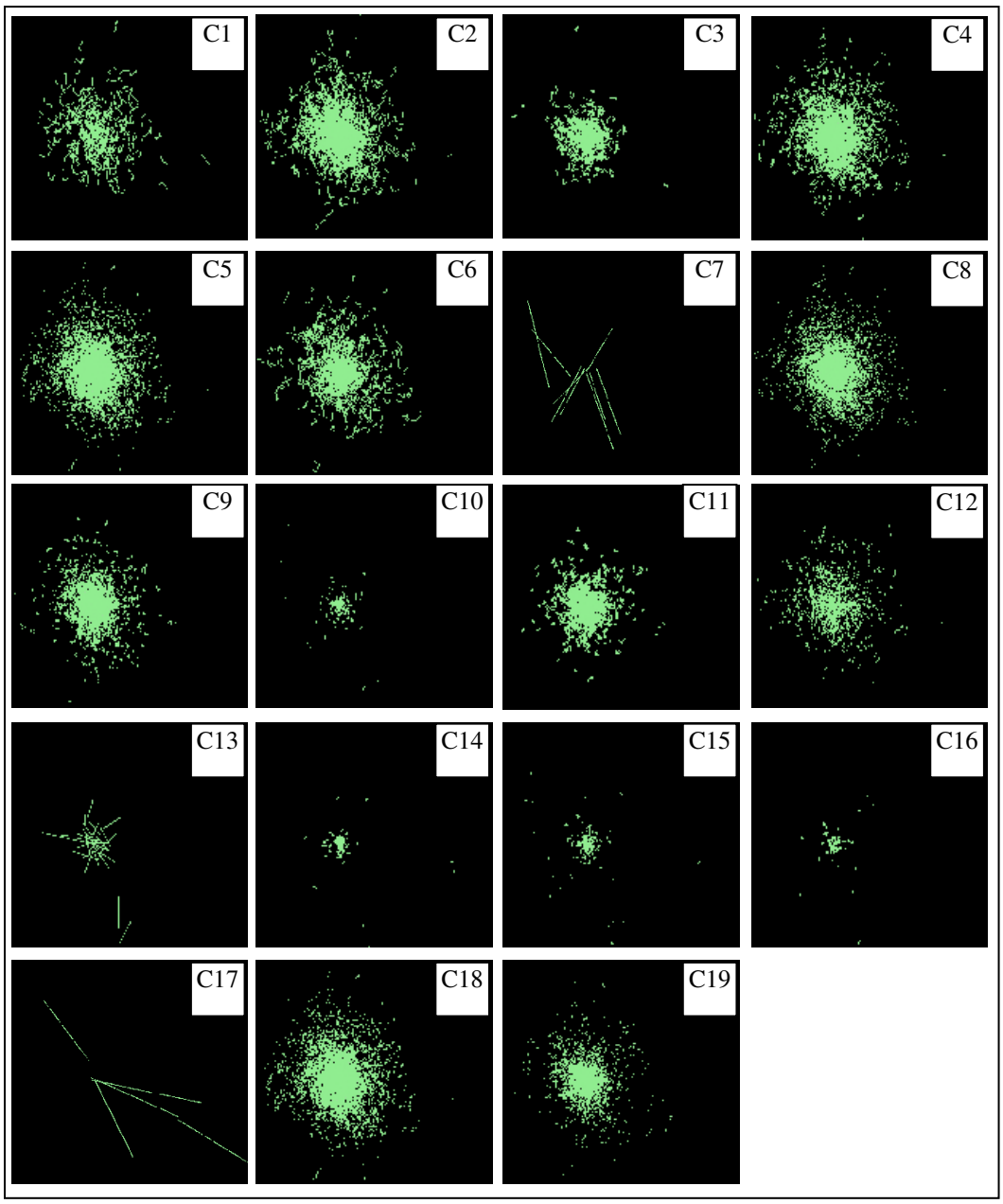

Figure 61. Sub-trajectory clusters (Distance-Threshold: *k*=19).

Figure 62 to Figure 64 are maps of cluster distribution through time. The vertical axis represents each trajectory ID from three random walk models, where trajectory IDs that equal to 1 to 100 are BM, 101 to 200 are CRW, and 201 to 300

are Lévy flight respectively. The horizontal axis is the simulation time. Each pixel in the images represents a cluster ID at a certain simulation time step for a trajectory. These maps allow us to visually recognize similarity and dissimilarity of trajectory clustering patterns through simulation time. In Figure 62, because trajectories are not partitioned, each trajectory is assigned by one cluster ID and thus one color throughout simulation. The images show that behaviors of three random walk models are roughly classified; however, it is clear that clustering by a set of whole trajectories with aggregated motion descriptors introduces misclassification particularly in trajectories between BM and CRW. The misclassification resulted from the similarity of global movement behaviors described by multiple motion descriptors. Because CRW is a probability model, some resultant trajectories can be more dispersed (Figure 65: Left image) while others can be more concentrated (Figure 65: Middle image). When looking at only global descriptors, those concentrated trajectories in CRW are more similar to the trajectories of BM (Figure 65: Right image). For example, the global sinuosity of such trajectories explained by straightness index and minimum boundary box is similar to that of trajectories in BM.

To avoid the confusion by global movement descriptors, trajectory partitioning approaches consider local behaviors. Figure 63 and Figure 64 show the temporal distribution of sub-trajectory clusters by TRACLUS-MDL and Distance-Threshold partitioning respectively. Both images illustrate that different random walk models share the same local behaviors, but the composition of those

behaviors are very different; therefore, two crisp boundaries that distinguish three random walks can be visually identified in each image.

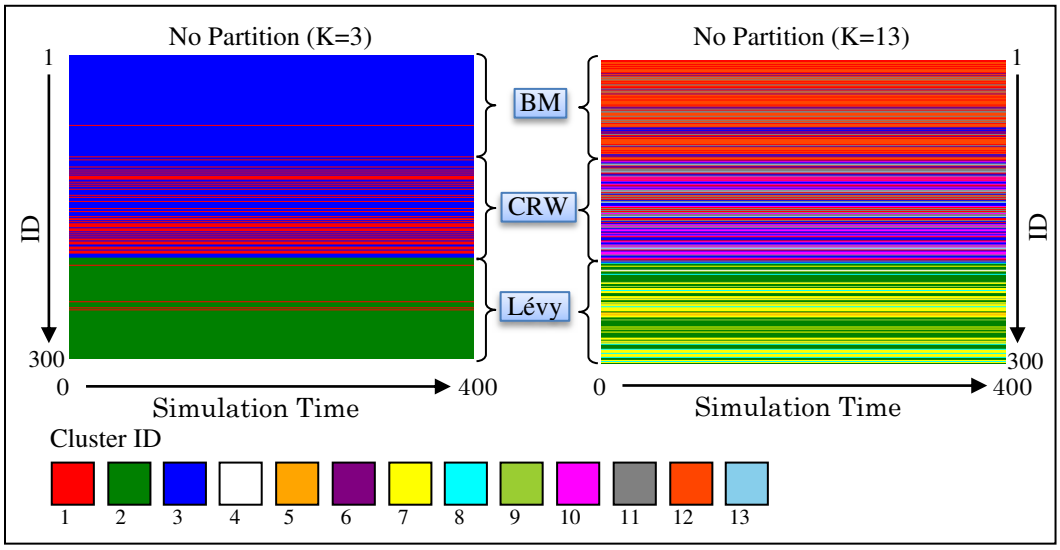

Figure 62. Temporal cluster distribution (no partition).

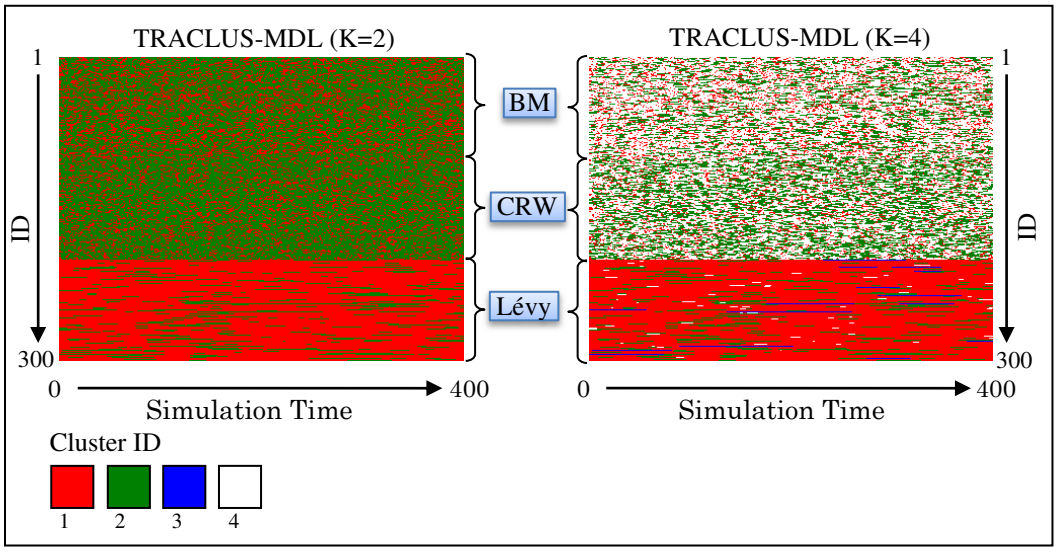

Figure 63. Temporal cluster distribution (TRACLUS-MDL).

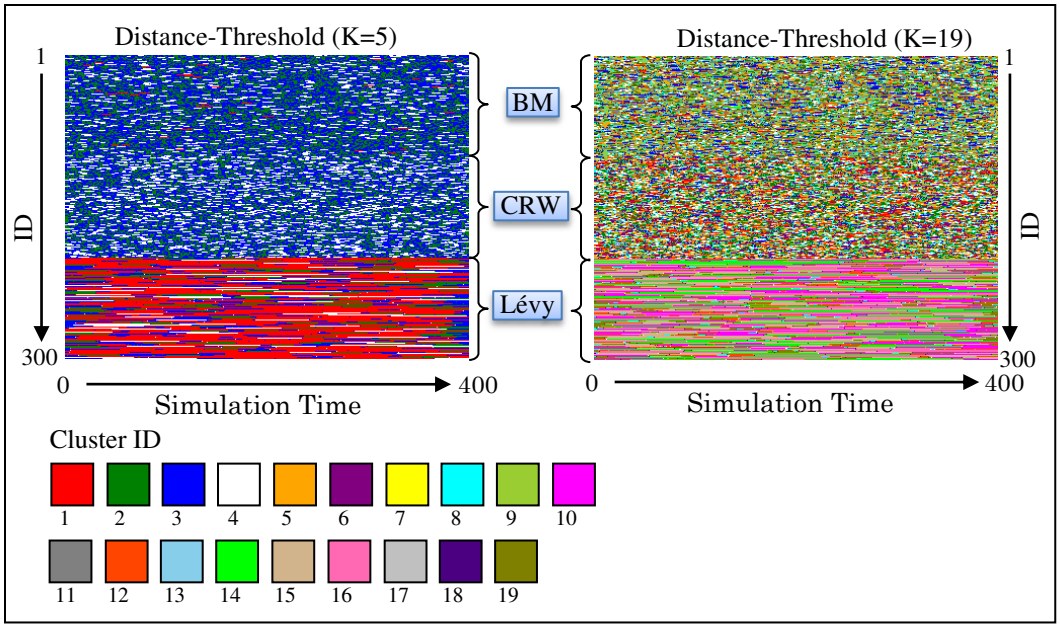

Figure 64. Temporal cluster distribution (Distance-Threshold).

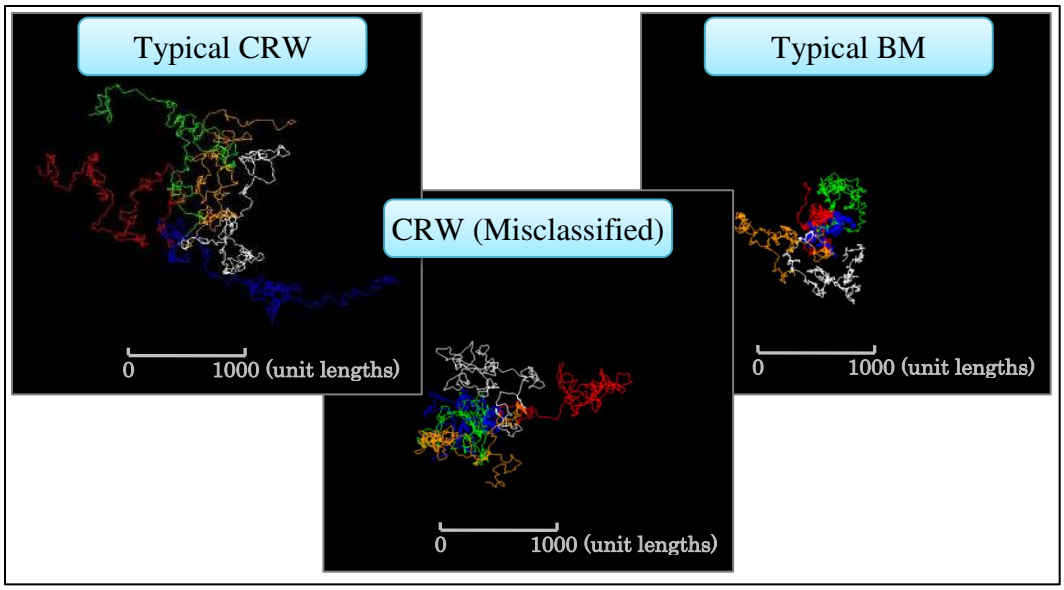

Figure 65. Five samples of misclassified trajectories in CRW using the no partitioning approach.

To quantitatively confirm how these cluster distributions can explain three random walk behaviors (BM, CRW, Lévy), J48, a Decision Tree algorithm, was applied. This is achieved by finding key behaviors (described by sub-trajectory clusters) that distinguish different random walk behaviors. The input dataset for the model was the total simulation time by each cluster in each trajectory. In J48, a parameter of confidence controls a pruning process, the smaller values of which incur more pruning. In this study, the value was set to 0.3. To evaluate the Decision Tree model for classification prediction, 10-folds cross validation was used. In the process, 90% of input data was used for training and 10% for testing, and the fold test was repeated 10 times, in which each set of data was used for testing once.

Table 12 presents the result of Decision Tree with a kappa coefficient that measures the agreement of prediction with the true class, where 1.0 signifies complete agreement between predicted and observed classes (Witten & Sander, 1981).

Table 13 lists confusion matrices for three partitioning methods. Because the numbers of k values for each partition algorithm are automatically determined by the gap statistics and they are different: the results cannot be compared directly. However, the recognition accuracy and kappa coefficient both show that the behavioral recognition of three random walks by trajectory partitioning algorithms is much higher than that of no-partition. This conforms that the compositions of sub-trajectory characteristics can better explain the global context of a trajectory than just aggregated characteristics of a whole trajectory. Between

partitioning algorithms, the Distance Threshold approach has higher accuracy with larger *k* values. TRUCLES-MDL has lower accuracy with lower *k* values. Again, even though the accuracy cannot be directly compared because of different *k* values, the theoretical principle of Occam's razor suggests that a better model is one which can explain the same phenomena with a lesser number of intellectual constructs (Batty & Torrens, 2005); thus, TRACLUS-MDL can be considered as a better algorithm to classify mixed random walk behaviors with less number of clusters. In order to directly compare the recognition accuracy between two partitioning algorithms, it is necessary to use sub-trajectory clusters with fixed number of *k*; however, the fixed *k* value will be no longer optimally determined by the gap statistics. The Occam's razor principle can be also applied to the number of *k* values; therefore, the smallest value of *k* determined by the gap statistic provides a better model for each partitioning approach, although the highest *k* values can produce higher accuracy of behavioral recognition.

Figure 66 shows a tree visualization of the Decision Tree result from the Distance Threshold approach with *k*=5. It shows key clusters that describe movement behaviors of three random walks. Sub-trajectories with Cluster 1 have negative values in *sct* and large duration in sub-trajectories (Figure 54 and Figure 60) suggesting staying behavior (i.e., small movements that are less than the distance threshold) with large turns. This behavior explains Lévy flight behavior by identifying a trajectory containing sub-trajectories of Cluster 1 more than 22 simulation time. Likewise, BM and CRW are distinguished by Cluster 4, sub-

trajectories of which have relatively longer and directed travel path (Figure 54 and Figure 60).

Figure 67 to Figure 74 illustrate sub-trajectory cluster distribution in a Space-Time Cube with two map representations, STPs and Space-Time line density. These maps visually confirm movement behaviors in three random walk models and the distribution of trajectory clustering results in space and time. Figure 67 to Figure 69 illustrate STPs colored by Cluster IDs in the Distance Threshold approach (k=5) for BM, CRW, and Lévy flight respectively. STPs in these maps can capture spatio-temporal movement behaviors in three random walks. In addition, STPs colored by Cluster IDs are useful in viewing spatio-temporal distributions of different movement behaviors. The comparison between Figure 67 and Figure 68 clarifies the difference in movements between BM and CRW, in which CRW are more dispersed from the origin point because of their directional correlation. Moreover, the difference of the composition of sub-trajectory clusters is visualized, where BM has several red spots that represent staying behavior (Cluster1) that do not appear in CRW. In addition, BM is more greenish drawn by Cluster 2, which implies sinuous walks, while CRW is more whitish drawn by Cluster 4, which implies directed movement. STPs of Lévy flight show significant difference from those of BM and CRW because the step length in Lévy flight follows a power distribution (Figure 69). The composition of sub-trajectory clusters is also significantly different. Trajectories of Lévy flight are composed of staying behavior (Cluster 1), two sinuous walks (Cluster 2 and 3), directed path (Cluster 4), and a very long directed path (Cluster 5).

Figure 70 to Figure 74 illustrate Space-Time line density maps (unit: unit lengths × unit area^{-1} × unit time^{-1}) of corresponding Cluster IDs (output voxel grid size: 200 × 200 (unit length) × 200 (unit time), bandwidth of STKDE: h_1=300 (unit length), h_2=300 (unit time)). Because all random walkers have the same origin, cluster density maps show higher values around the origin and they are dispersed as simulation time elapsed. In addition, because of random walk models, each cluster is randomly distributed in the Space-Time Cube.

Table 12. Results of decision tree classification.

Partition Algorithm	k	Classification			
		Corr.	Incorr.	Corr. (%)	Kappa
No Partition	3	238	62	79.33	0.69
	13	249	51	83.00	0.74
TRACLUS-MDL	2	266	34	88.67	0.83
	4	280	20	93.33	0.90
Distance Threshold	5	267	33	89.00	0.84
	19	296	4	98.67	0.98

Table 13. Confusion matrix of behavioral recognition.

No partition					TRACLUS-MDL					Distance-Threshold				
K		Classified As			K		Classified As			K		Classified As		
		BM	CRW	Lévy			BM	CRW	Lévy			BM	CRW	Lévy
3	BM	99	1	0	2	BM	84	15	1	5	BM	90	8	2
	CRW	56	44	1		CRW	17	83	1		CRW	21	79	1
	Lévy	0	4	95		Lévy	0	0	99		Lévy	0	1	98
13	BM	71	29	0	4	BM	89	10	1	19	BM	100	0	0
	CRW	18	82	1		CRW	8	92	1		CRW	3	97	1
	Lévy	0	3	96		Lévy	0	0	99		Lévy	0	0	99

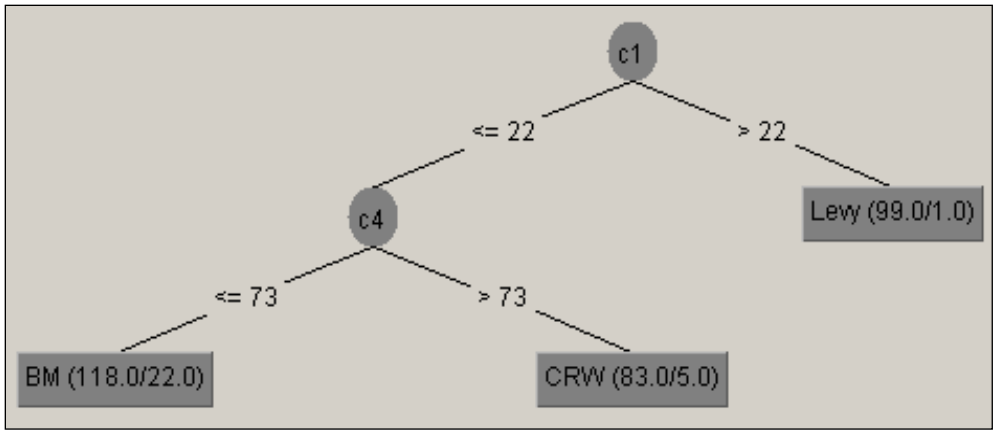

Figure 66. A tree visualization of Decision Tree results (Distance-Threshold: *k*=5).

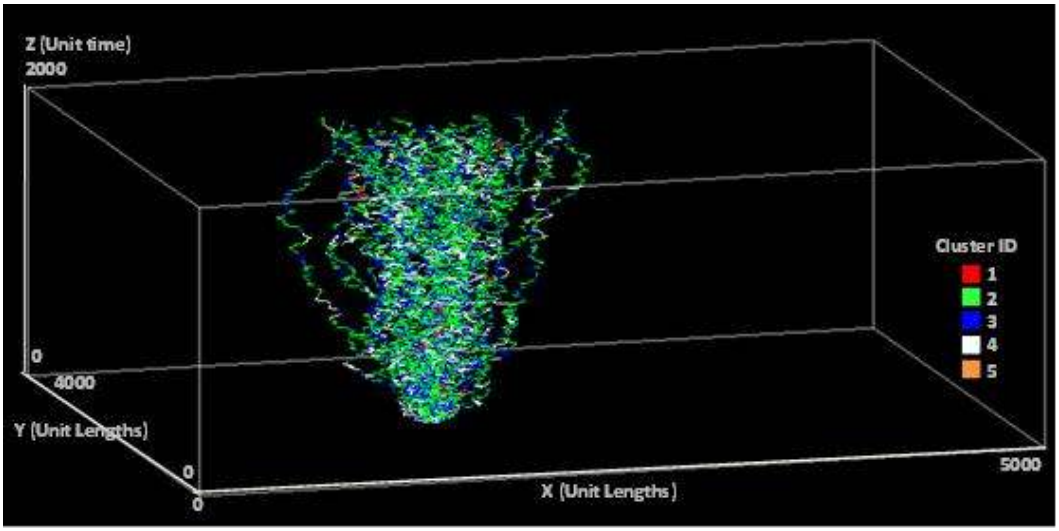

Figure 67. STPs of BM colored by cluster ID.

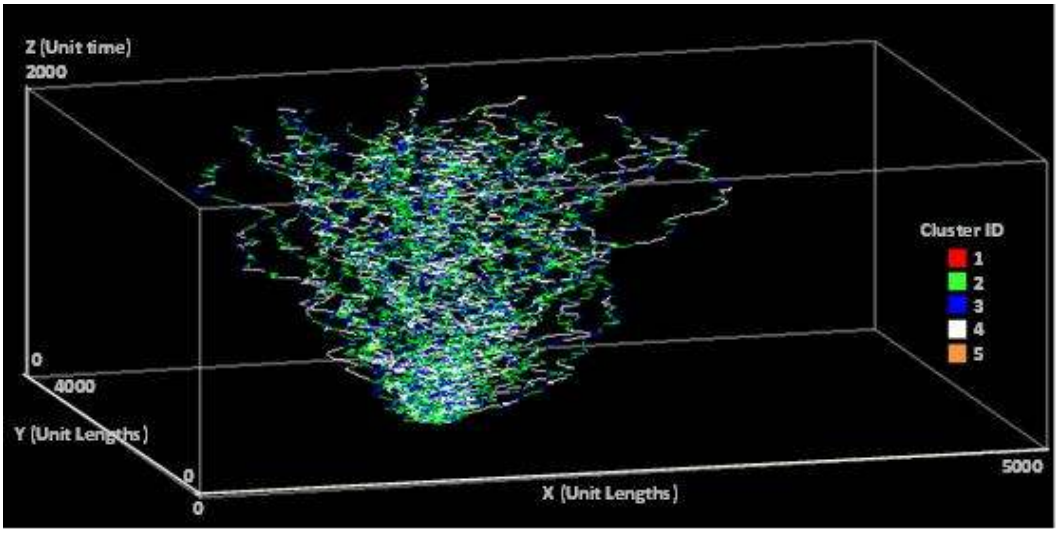

Figure 68. STPs of CRW colored by cluster ID.

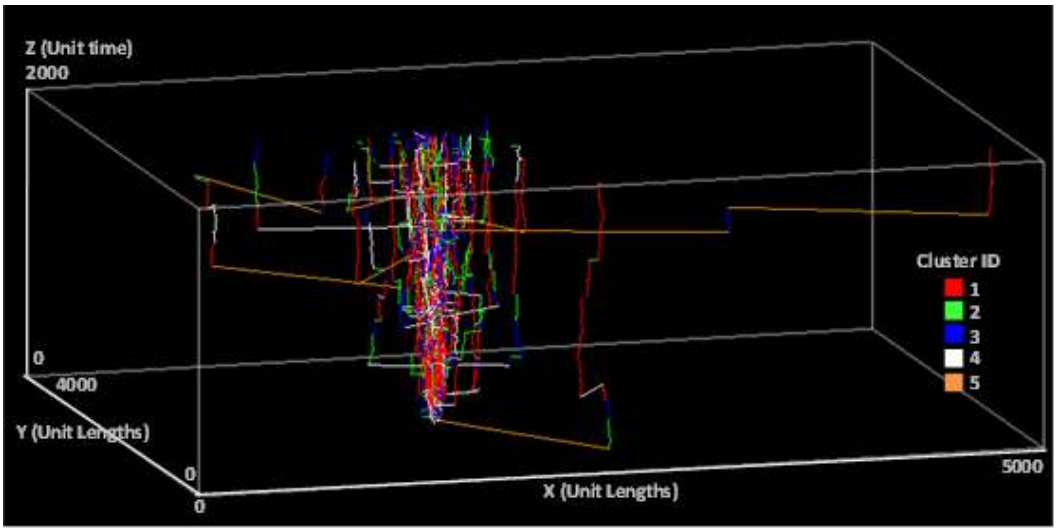

Figure 69. STPs of Lévy Flight colored by cluster ID.

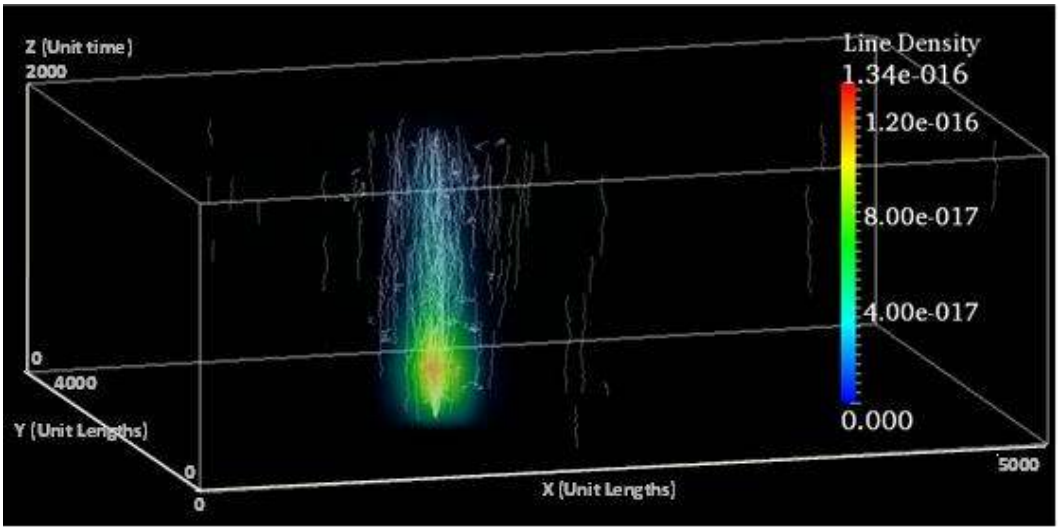

Figure 70. Space-Time line density map of Cluster 1.

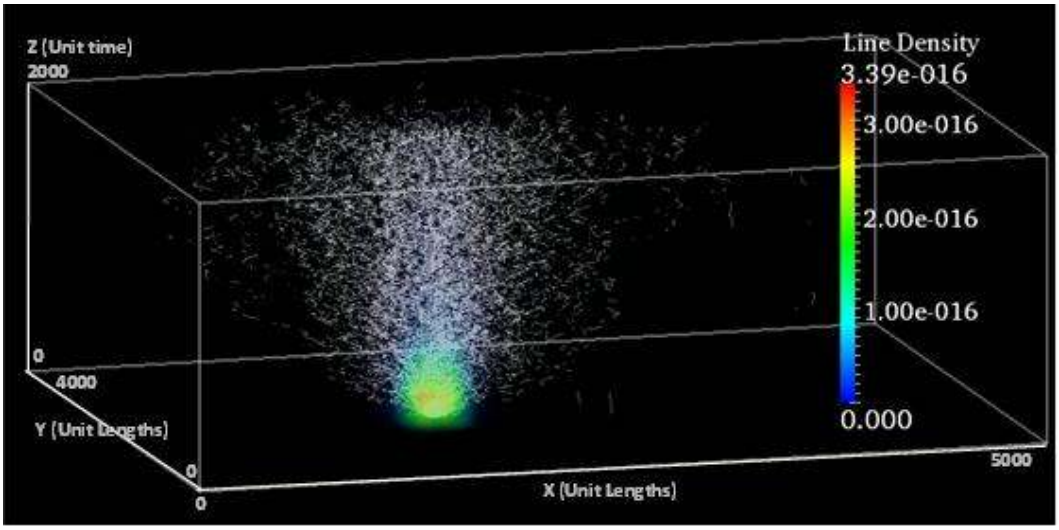

Figure 71. Space-Time line density map of Cluster 2.

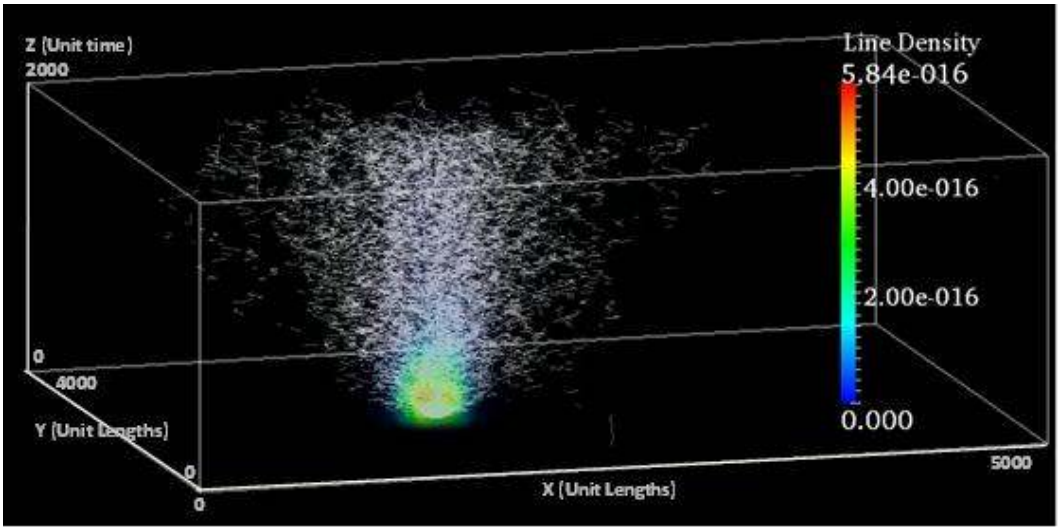

Figure 72. Space-Time line density map of Cluster 3.

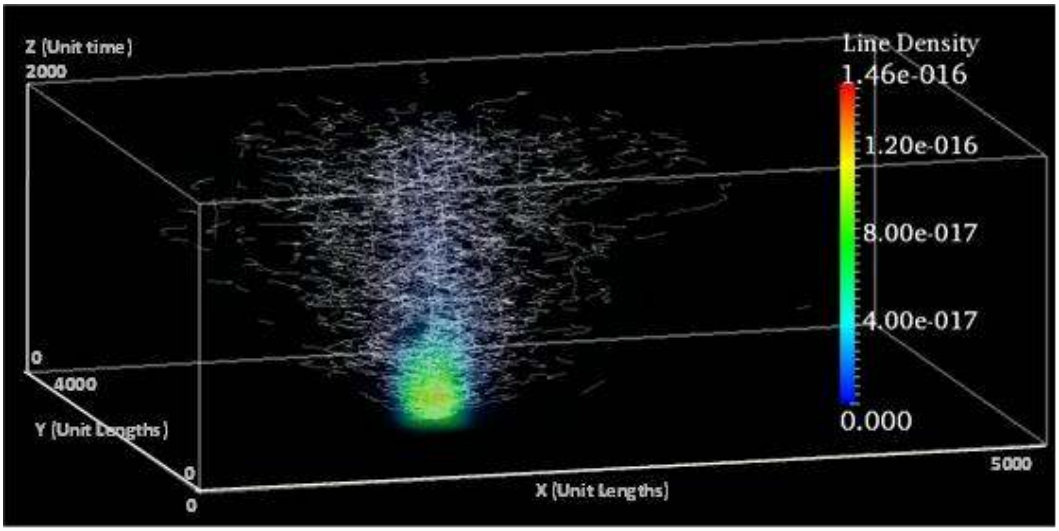

Figure 73. Space-Time line density map of Cluster 4.

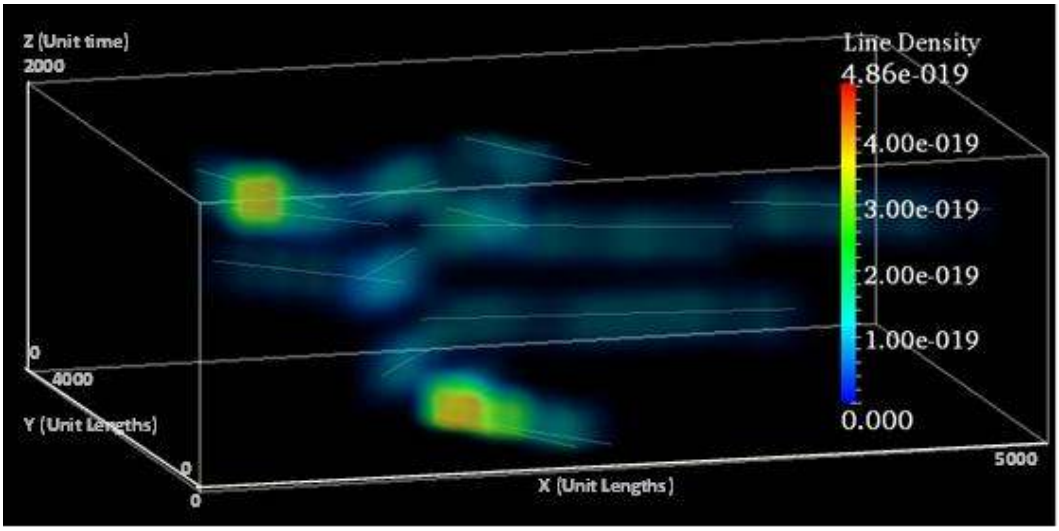

Figure 74. Space-Time line density map of Cluster 5.

5.4.2 Trajectory Data-Mining on GPS Data

It is also useful to test the data-mining scheme for real-world movement, which may be more "organic" than data generated in simulation. Ideally, it would be useful to build a scheme that can mine and compare both real and simulation data. With this in mind, I ran the scheme over GPS tracks of real movement.

5.4.2.1 Dataset

In this study, GPS logs were collected for one subject's daily movement using the GARMIN GPSMAP 60CS GPS receiver, a 12-parallel-channel receiver that continuously tracks and uses up to 12 satellites to compute and update position information. The sampling frequency was set to one point per second. To increase the accuracy of measurements, communication with MSAS (Multi-functional Transport Satellite), a Japanese WAAS (Wide Area Augmentation System), was

enabled. With MSAS, the GPS accuracy on position can increase from < 15 meters (95% typical) to < 3 meters (95% typical). The dataset was collected in the Kansai area, Japan, between April and June 2010, providing GPS logs for 36 days (371.68 hours) with 335.575 points. I also applied a resampling procedure that generates a trajectory at regular intervals by linear interpolation along the original trajectory in order to reduce the data size and to insert points where no data were recorded. In this study, the raw GPS logs with sampling frequency at 1 second were resampled by 30 seconds intervals. Figure 75 displays the entire study area and GPS trajectories, while Figure 76 shows on an enlarged view around the area of the subject's residence.

Along with GPS logs, behavioral contexts of major daily activities and transportation modes were collected. There are five major activities; "WORK", "DINING" as dining-out activities, "TRIP" as traveling across prefectures, "SHOPPING" that excludes daily grocery or commodity shopping, and "EXERCISE" such as walking and jogging. To match activities to trips, a trajectory of a single day was hand labeled with one major activity or two if there was another major activity observed. In addition, the major activity label was further aggregated into a binary activity label, "WORK" and "Non-Work". These labels were used for evaluating the unsupervised learning of the trajectory data-mining framework.

Table 14 presents frequency of major and binary daily activities from 36 days samples. Transportation modes recorded include six types: "Walk", "Run", "Train (local)", "Train (express)", "Subway", "Bus", "Light rail", and "Car".

It is important to note that a trajectory dataset collected by LATs contains uncertainty. For example, various measurement errors can exist in a dataset due to the quality of device, environment factors (e.g., existence of obstacles blocking signals, multi-path effects by signal reflection), and human oriented errors (e.g., missing values due to device inactivity, wrong positioning by leaving the device at home). As another example, a common approach of linear interpolation for the resampling method relies on the (unrealistic) assumption that between two sample points, an object unidirectionally moves at constant speed. This study uses datasets that include potential uncertainties. Nevertheless, because the proposed framework of trajectory data-mining identifies local behavioral patterns of movement, it could detect a cluster of sub-trajectories that associates with above mentioned uncertainties.

Table 14. Frequency of activities.

Major Activity	Frequency	Binary Activity	Frequency
Work	22	Work	27
Dining	1	Non-Work	9
Trip	5		
Exercise	1		
Work&Dining	4		
Shopping&Dining	2		
Trip&Work	1		
Total	36		36

Figure 75. Study area and GPS trajectories.

Figure 76. GPS trajectories in the area of a subject's residence.

5.4.2.2 Results

Similar to the previous experiment, in order to evaluate the effect of different trajectory partitioning approaches in trajectory data-mining, three partitioning

algorithms were compared: no-partitioning, TRACLUS-MDL, and Distance-Threshold. For trajectory partitioning, the parameter value of c was set to 0.3 in the TRACLUS-MDL approach, while Th_d, was set to 5 meters in the Distance-Threshold approach. Figure 77 and Figure 78 represent the results of trajectory partitioning for GPS tracking data by two partitioning algorithms with two-tone STP coloring. In the figures, each partitioned Space-Time trajectory in a trajectory is alternately colored by red and cyan. In the TRACLUS-MDL approach, a trajectory is partitioned where the geometrical shape is suddenly changed (i.e., large directional changes). In the Distance-Threshold approach, a trajectory is partitioned where a segment is less than the defined threshold value indicating very slow movement or staying behavior.

The STP maps show that both approaches partitioned a trajectory at long staying behaviors, represented as vertical lines in the Space-Time Cube. While the Distance Threshold approach preserves a long segment of 2D movement between staying segments, the TRACLUS-MDL approach has fragmented segments partitioned by geometrical changes. In terms of human movements, a long segment partitioned by two staying points using Distance-Threshold partitioning could contain multiple continuous movement behaviors such as walking and running when a person exercises. On the other hand, one continuous movement behavior such driving a car on a high-way or taking an express train may have multiple curves, but such a path will be fragmented by TRACLUS-MDL partitioning.

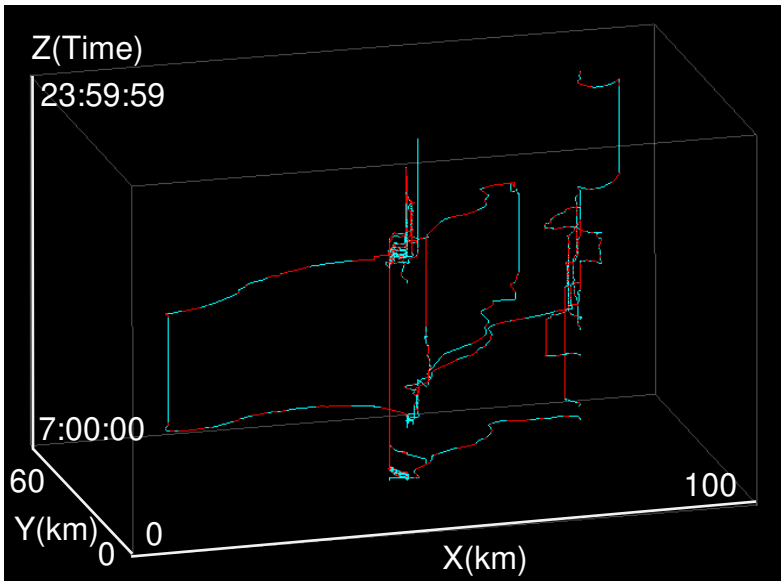

Figure 77. Two-tone STP representation of trajectory partitioning (TRACLUS-MDL).

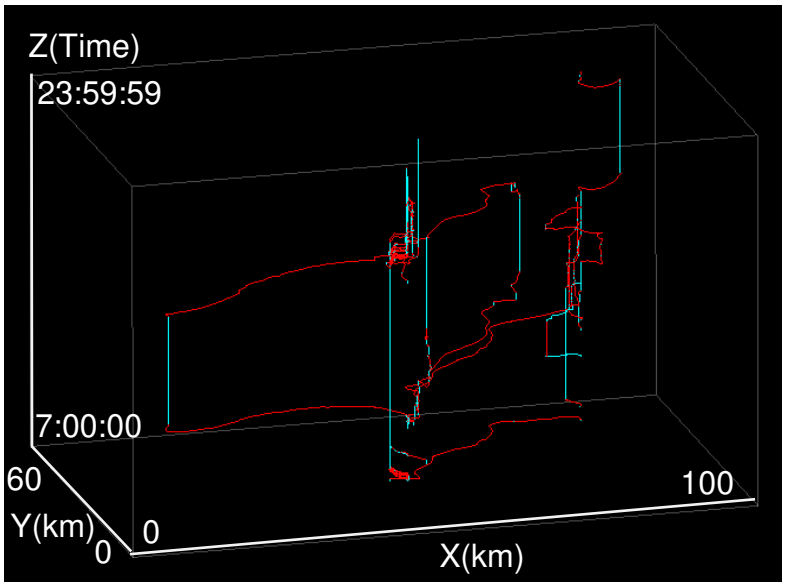

Figure 78. Two-tone STP representation of trajectory partitioning (Distance Threshold).

For each trajectory partition in three different partition algorithms, multi-dimensional vectors were calculated to characterize the partition trajectory. To identify dependencies of multiple motion variables in each sub-trajectory dataset, correlation analysis was performed. Tables Table 15 to Table 17 list correlation matrices for movement variables of trajectory partition for the three partition algorithms. Between TRACLUS-MDL and Distance-Threshold, the structure of the correlation matrix is similar and no large difference is found. This implies that variable relationships in two partitioned datasets are similar. Between no-partition and the two partitioning approaches, large differences are found in the correlation between distances (d_x, d_y, d_{2D}) and beeline distances (d_{sx}, d_{sx}, d_{s2D}), where the no partition approach has positive correlation around 0.4 and two partitioning approaches both have very high positive correlation around 0.9. The relationship between distance and beeline distance describes sinuosity of a path. The positive correlation explains that a path is more likely to be straight if the distance is longer. This is reasonable because the dataset used in this study contains long directed paths by train trips. Moreover, the relationship is stronger for sub-trajectory than for whole trajectory. This is also reasonable because a whole trajectory in this dataset is a daily trip and it may contain multiple behaviors, but a partitioned one may only have one behavior with long directed movement.

To reduce the dimensionality of the dataset consisting of interrelated variables, PCA was performed. Table 18 to Table 20 show the results of PCA for three partitioning algorithms. The numbers of identified Principal Components (PCs) with eigenvalue greater than 1 are 3, 2, and 2 for no-partition, TRACLUS-

MDL, and Distance-Threshold respectively. These PCs explain 95.4%, 75.1%, and 81.4% of original variables for the dataset in three partitioning approaches respectively. In no-partition, the first PC describes short travel length and slow movement, the second PC describes directed movement, and the third PC describes staying behavior. In TRACLUS-MDL, the first PC describes short travel length and slow movement, and the second PC describes sinuous movement. In Distance-Threshold, the first PC describes short and slow and sinuous movement, and the second PC describes fast movement.

Table 15. Correlation matrix for movement variables (GPS: no partition).

	dt	*dx*	*dy*	*d2D*	*v*	*vx*	*vy*	*dsx*	*dsy*	*ds2D*	*mbb*	*sct*
dt	1											
dx	-0.0798	1										
dy	-0.0134	0.8945	1									
d2D	-0.0559	0.9780	0.9680	1								
v	-0.4902	0.8787	0.7922	0.8642	1							
vx	-0.5089	0.8630	0.6949	0.8099	0.9817	1						
vy	-0.4406	0.8396	0.8741	0.8803	0.9639	0.8965	1					
dsx	-0.0791	0.4026	0.4391	0.4283	0.4204	0.3647	0.4775	1				
dsy	-0.0570	0.4060	0.4429	0.4320	0.4085	0.3514	0.4666	0.9963	1			
ds2D	-0.0726	0.4037	0.4406	0.4296	0.4170	0.3607	0.4745	0.9995	0.9985	1		
mbb	-0.0453	0.7853	0.8529	0.8366	0.7185	0.6365	0.7897	0.8172	0.8159	0.8166	1	
sct	0.0139	0.7666	0.8543	0.8288	0.6510	0.5634	0.7244	0.2365	0.2441	0.2385	0.6059	1

Table 16. Correlation matrix for movement variables (GPS: TRACLUS-MDL).

	dt	*dx*	*dy*	*d2D*	*v*	*vx*	*vy*	*dsx*	*dsy*	*ds2D*	*mbb*	*sct*
dt	1											
dx	-0.0018	1										
dy	0.0019	0.6735	1									
d2D	-0.0002	0.9336	0.8850	1								
v	-0.0407	0.6842	0.6547	0.7363	1							
vx	-0.0367	0.7361	0.4834	0.6809	0.9196	1						
vy	-0.0362	0.4809	0.7446	0.6482	0.8732	0.6261	1					
dsx	-0.0043	0.9947	0.6660	0.9253	0.6886	0.7422	0.4833	1				
dsy	0.0000	0.6656	0.9912	0.8744	0.6619	0.4877	0.7551	0.6664	1			
ds2D	-0.0026	0.9304	0.8776	0.9934	0.7440	0.6882	0.6550	0.9326	0.8803	1		
mbb	-0.0025	0.7294	0.7957	0.7958	0.4982	0.4566	0.4877	0.7366	0.8054	0.806	1	
sct	-0.0058	0.4019	0.3478	0.4087	0.1686	0.1622	0.1478	0.4032	0.3493	0.4116	0.2616	1

Table 17. Correlation matrix for movement variables (GPS: Distance-Threshold).

	dt	*dx*	*dy*	*d2D*	*v*	*vx*	*vy*	*dsx*	*dsy*	*ds2D*	*mbb*	*sct*
dt	1											
dx	0.0132	1										
dy	0.0116	0.8895	1									
d2D	0.0127	0.9795	0.9629	1								
v	-0.0316	0.5014	0.5091	0.5218	1							
vx	-0.0292	0.5970	0.5179	0.5786	0.8772	1						
vy	-0.0282	0.3595	0.4412	0.4085	0.9450	0.6913	1					
dsx	0.0136	0.9954	0.8698	0.9682	0.4961	0.5942	0.3523	1				
dsy	0.0119	0.8458	0.9843	0.9308	0.4847	0.4806	0.4300	0.8278	1			
ds2D	0.0133	0.9793	0.9533	0.9958	0.5209	0.5775	0.4070	0.9744	0.9315	1		
mbb	0.0180	0.8753	0.9539	0.9346	0.3868	0.4145	0.3202	0.8659	0.9685	0.9395	1	
sct	-0.0037	0.6673	0.6246	0.6651	0.3499	0.4159	0.2550	0.6487	0.5483	0.6410	0.4981	1

Table 18. Results of PCA (GPS: no partition).

Variables	Principal Components			
	Loadings			Contribution
	1	2	3	
d_t	0.2354	0.2280	0.9137	0.9422
d_x	-0.9099	-0.2711	0.1739	0.9317
d_y	-0.9085	-0.1968	0.2996	0.9538
d_{2D}	-0.9340	-0.2475	0.2334	0.9881
v	-0.9085	-0.3031	-0.2589	0.9843
v_x	-0.8483	-0.3332	-0.3172	0.9313
v_y	-0.9353	-0.2356	-0.1724	0.9600
d_{sx}	-0.6783	0.7263	-0.0920	0.9961
d_{sy}	-0.6753	0.7304	-0.0669	0.9940
d_{s2D}	-0.6777	0.7283	-0.0848	0.9968
mbb	-0.9305	0.2841	0.1289	0.9632
sct	-0.7422	-0.3436	0.3736	0.8086
Eigen.values	7.7931	2.2741	1.3830	11.4502
Proportion	64.94	18.95	11.53	95.42
Cumulative.prop.	64.94	83.89	95.42	-

Table 19. Results of PCA (GPS: TRACLUS-MDL).

Variables	Principal Components		
	Loadings		Contribution
	1	2	
d_t	0.0147	0.2532	0.0643
d_x	-0.9045	0.1394	0.8375
d_y	-0.8922	0.0941	0.8048
d_{2D}	-0.9765	0.1215	0.9683
v	-0.8356	-0.4881	0.9364
v_x	-0.7652	-0.4278	0.7685
v_y	-0.7549	-0.4397	0.7632
d_{sx}	-0.9051	0.1348	0.8373
d_{sy}	-0.8936	0.0860	0.8059
d_{s2D}	-0.9802	0.1164	0.9743
mbb	-0.8149	0.2262	0.7152
sct	-0.4019	0.6094	0.5329
Eigen.values	7.8254	1.1833	9.0087
Proportion	65.21	9.86	75.07
Cumulative.prop.	65.21	75.07	-

Table 20. Results of PCA (GPS: Distance-Threshold).

Variables	Principal Components		
	Loadings		Contribution
	1	2	
d_t	-0.0043	-0.1046	0.0110
d_x	-0.9547	-0.1694	0.9402
d_y	-0.9555	-0.1648	0.9401
d_{2D}	-0.9823	-0.1697	0.9936
v	-0.6599	0.7464	0.9926
v_x	-0.6891	0.5869	0.8193
v_y	-0.5513	0.7631	0.8863
d_{sx}	-0.9445	-0.1677	0.9202
d_{sy}	-0.9279	-0.1765	0.8921
d_{s2D}	-0.9801	-0.1701	0.9896
mbb	-0.9080	-0.2911	0.9093
sct	-0.6780	-0.1152	0.4730
Eigen.values	8.0016	1.7656	9.7672
Proportion	66.68	14.71	81.39
Cumulative.prop.	66.68	81.39	-

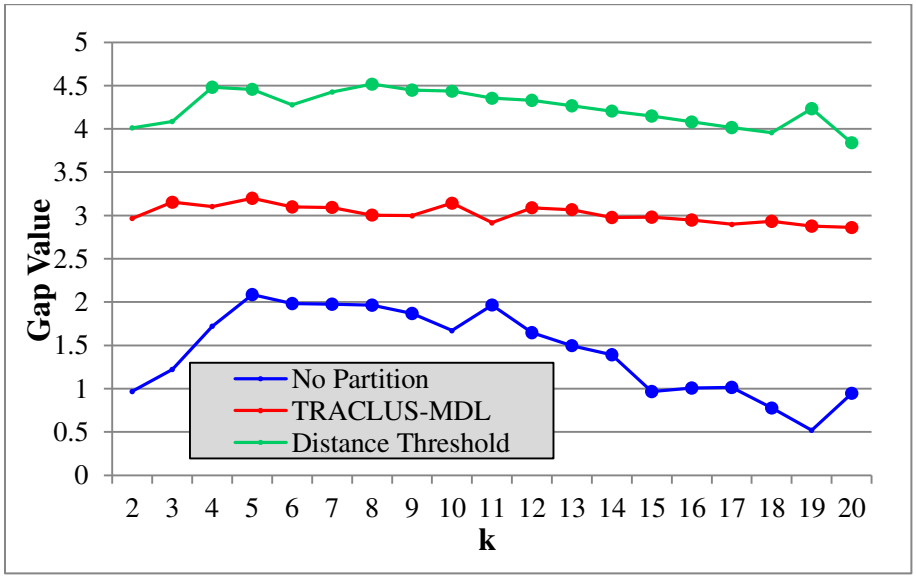

Figure 79. Gap curve for three partitioning algorithms (GPS).

PC scores of each sub-trajectory for each PC (Eigen value ≥ 1) were calculated, and then they were used as a new input dataset for cluster analysis. K-means clustering was run for each sub-trajectory dataset in three partition algorithms with different *k* in a range between 2 and 20, which is arbitrarily defined. The optimal values of $\hat{k}$ were estimated by applying the gap statistic, in which $\hat{k} = smallest\ k\ such\ that\ Gap(k) \geq Gap(k+1) - s(k+1)$. The number of generating reference datasets of a null model, *B*, was set to 25. Figure 79. illustrates gap curves for three partition algorithms, where large dots represent *Gap(k)* greater than or equal to *Gap(k+1) – s(k+1)*. As in the previous section, this study also considers the number of *k* determined by the highest value of *Gap(k)* in the range of *k* between 2 and 20 as an alternative value because the highest gap value represents the largest difference of the compactness of clusters between a raw dataset and a null reference dataset (i.e., random distribution in this study). Following results of cluster analysis and the gap statistics, optimal values of *k* are 5, 3, and 4 for no-partition, TRACLUS-MDL, and Distance-Threshold respectively. *k* values determined by the highest value of *Gap(k)* are 5, 5, and 8 respectively. Figure 80 to Figure 84 illustrate the numbers of sub-trajectories assigned to a cluster for corresponding partitioning methods and selected *k* values.

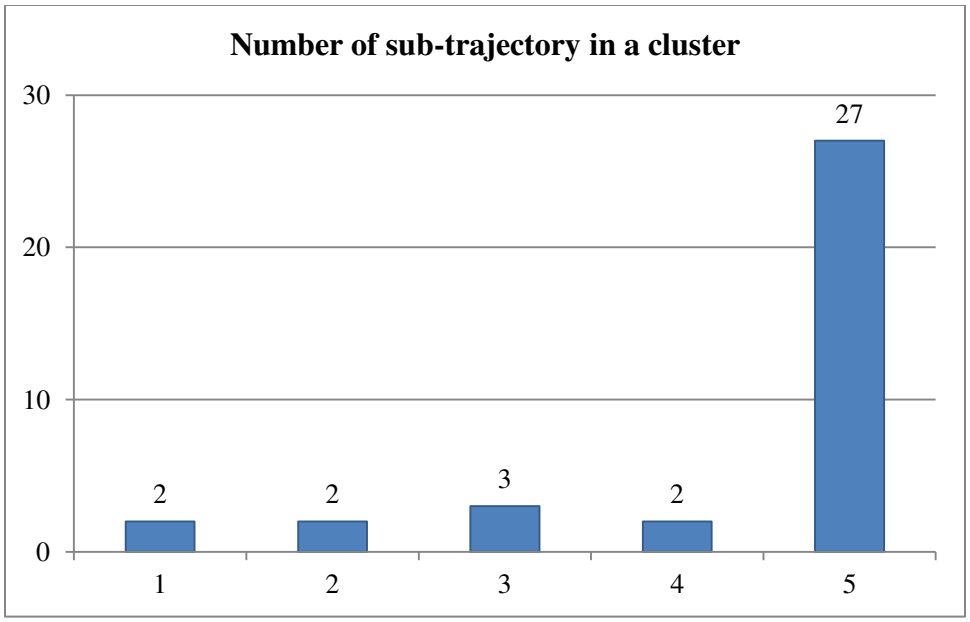

Figure 80. Number of subtrajectories in a cluster (n=36) (no partition: *k*=5).

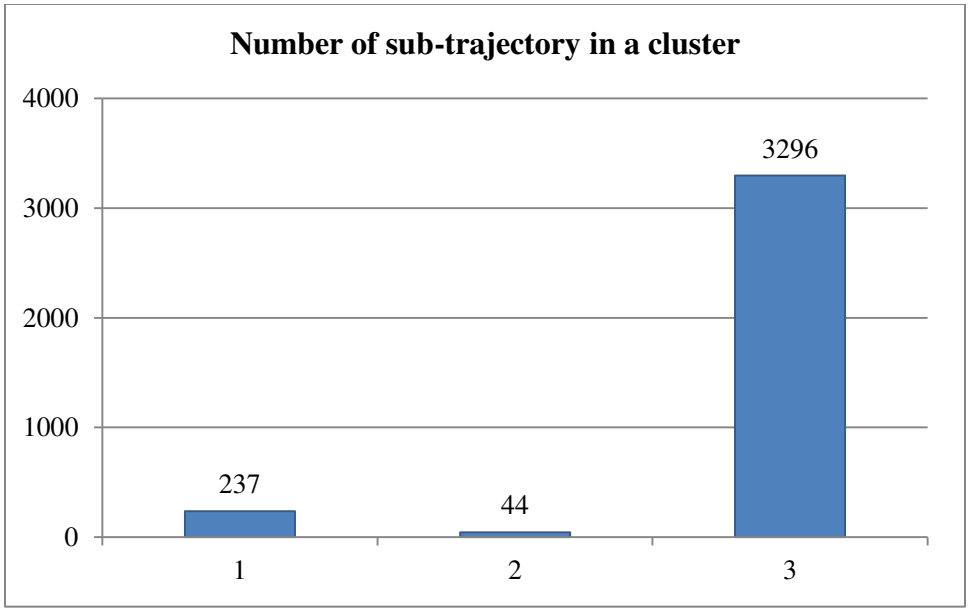

Figure 81. Number of subtrajectories in a cluster (n=36) (TRACULS-MDL: *k*=3).

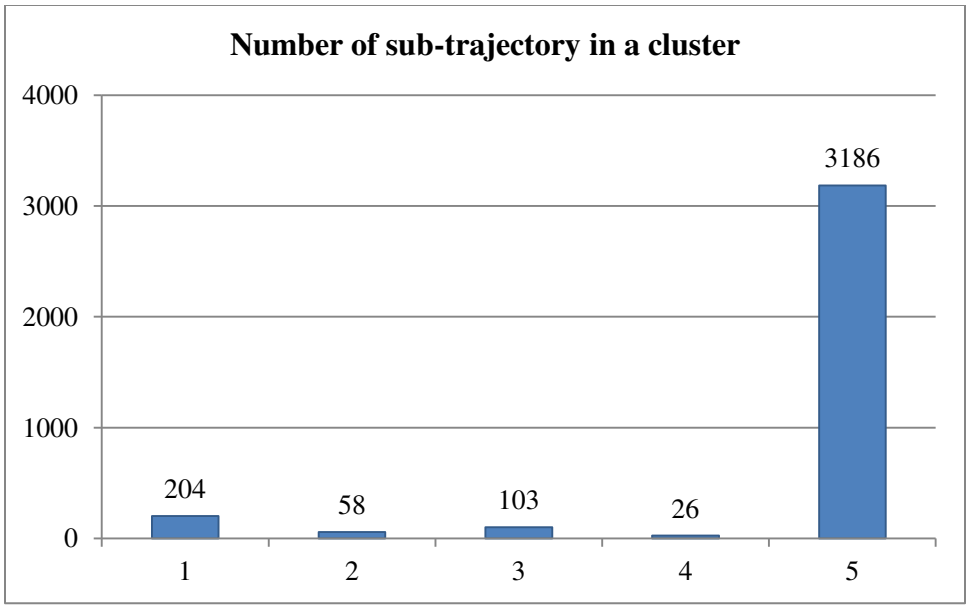

Figure 82. Number of subtrajectories in a cluster (n=36) (TRACULS-MDL: *k*=5).

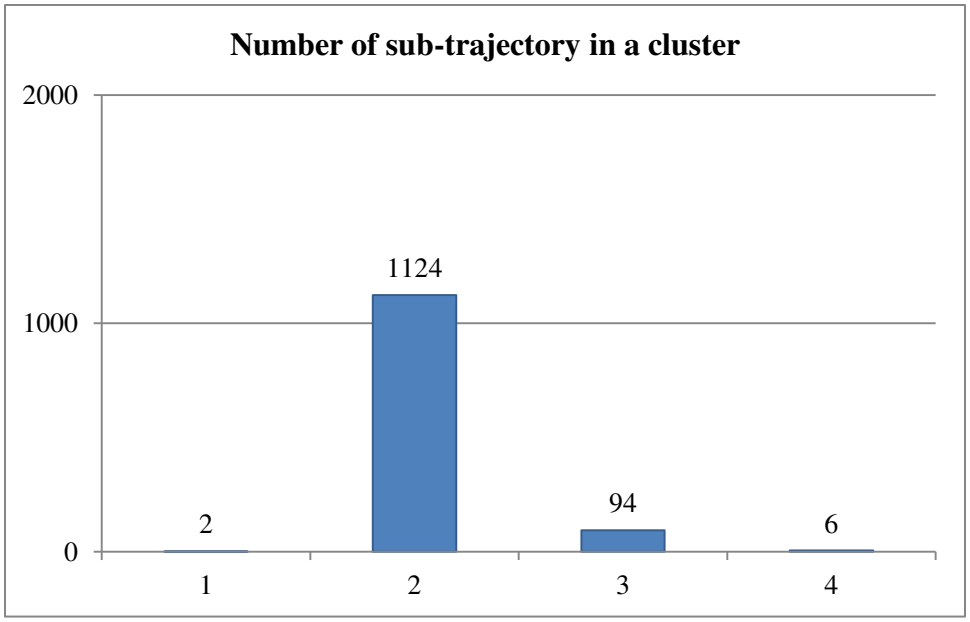

Figure 83. Number of subtrajectories in a cluster (n=36) (Distance-Threshold: *k*=4).

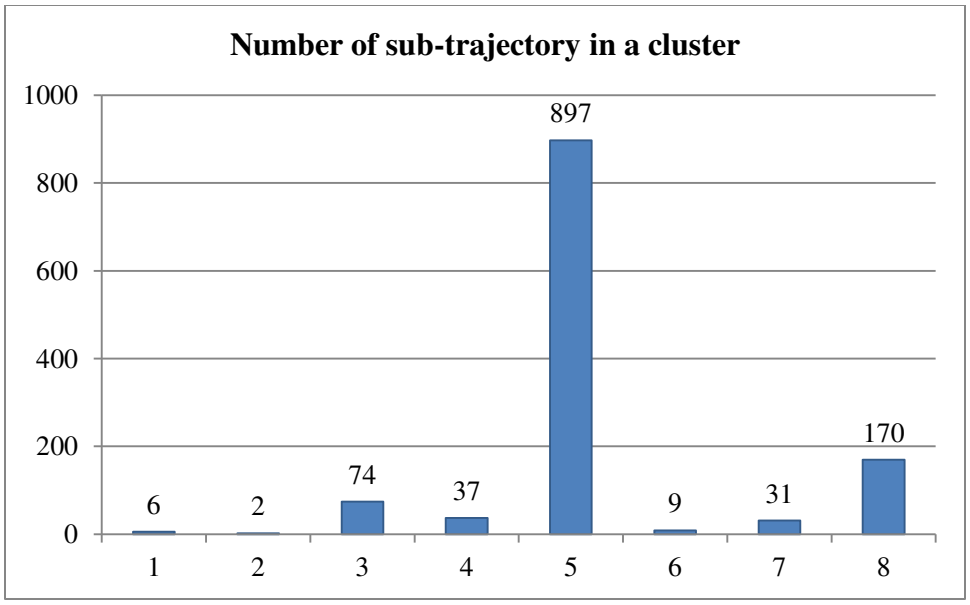

Figure 84. Number of subtrajectories in a cluster (n=36) (Distance-Threshold: *k*=8).

Figure 85 to Figure 89 present cluster profiles for corresponding partitioning methods and selected *k* values, where the vertical axis is cluster ID and the horizontal axis shows the average of normalized value of independent variables within a cluster. Figure 90 to Figure 94 display sub-trajectories for each cluster ID for corresponding partitioning methods and selected *k* values. These figures explain sub-trajectory characteristics within a cluster.

In the no partition algorithm (which treats a trajectory as a whole), the optimal *k* value estimated by the gap statistic and the highest gap value was 5. The cluster profile (Figure 85) and the image of trajectories (Figure 90) describe the following behaviors; Cluster 1 represents relatively moderate duration, moderate travel length, moderate velocity, and sinuous path (small beeline distance) with

smooth turns (large *sct*). According to the travel diary, the major daily activity of Cluster 1 was labeled as “Trip” and the transportation mode of the trip was “Car”. Cluster 2 shows moderate duration, large travel distance, high velocity, and directed movement (large beeline distance) with smooth turns. The travel diary showed the major activity of these trajectories as “Trip” and the transportation mode of the trip was “Train”. Cluster 3 represents very short duration, short travel distance, moderate velocity, and sinuous movement. Major activities of these trajectories included “Trip” and “Exercise”. Cluster 4 represents moderate duration, long travel distance, high velocity, and sinuous movement with smooth turns, where major activities were labeled as “Trip” and the transportation mode of the trip was “Train”. Cluster 5 represents large duration, short travel distance, low velocity, and sinuous movement. This explains staying behaviors and the major activity of Cluster 5 was labeled as “Work”. Because the no-partitioning approach uses an entire trajectory, the clustering result generally matches major activities in the travel diary; however, the global approach cannot capture local behaviors during a single day.

In TRACLUS-MDL, optimal k value was 3 estimated by the gap statistics and 5 estimated by the highest gap value. Trajectory clusters in this approach describe local movement behaviors. When $k = 3$, Cluster 1 represents relatively moderate travel distance and velocity such as trips by bus; Cluster 2 represents long travel distance, high velocity, and directed movement such as trips by train; and Cluster 3 represents short travel distance, low velocity, and sinuous movement such as staying behaviors. When $k = 5$, the local movement behavior

of Cluster 1 when $k = 3$ is further divided into 3 clusters (Cluster 1, 2, and 3 with $k = 5$).

In Distance-Threshold, the optimal k value estimated by the gap statistic was 4 and estimated by the highest value of gap statistic was 8. When $k = 4$, Cluster 1 represents short travel distance and very high velocity explaining irregular paths; Cluster 2 represents short travel distance with slow movement such as walks and working at an office; Cluster 3 represents moderate travel distance with moderate velocity such as trips by bus and car; and Cluster 4 represents long travel distance, high velocity, and directive movement such as trips by train. When $k = 8$, Cluster 1 corresponds to train trips (Cluster 4 for $k = 4$); Cluster 2 corresponds to irregular paths (Cluster 1 for k = 4); Cluster 3, 4, and 6 correspond to car and bus trips (Cluster 3 for $k = 4$) with different degrees of travel distance, velocity, and sinuosity; Cluster 5, 7, and 8 correspond to Cluster 2 for $k = 4$ where Cluster 5 and 8 describe slow movements like walks and runs and Cluster 7 describes staying behaviors such as working at an office. The key difference between TRACLUS-MDL and Distance-Threshold is, again, the treatment of staying behavior. Distance-Threshold with $k = 8$ successfully extracted one cluster that explain only staying behavior (Cluster 7).

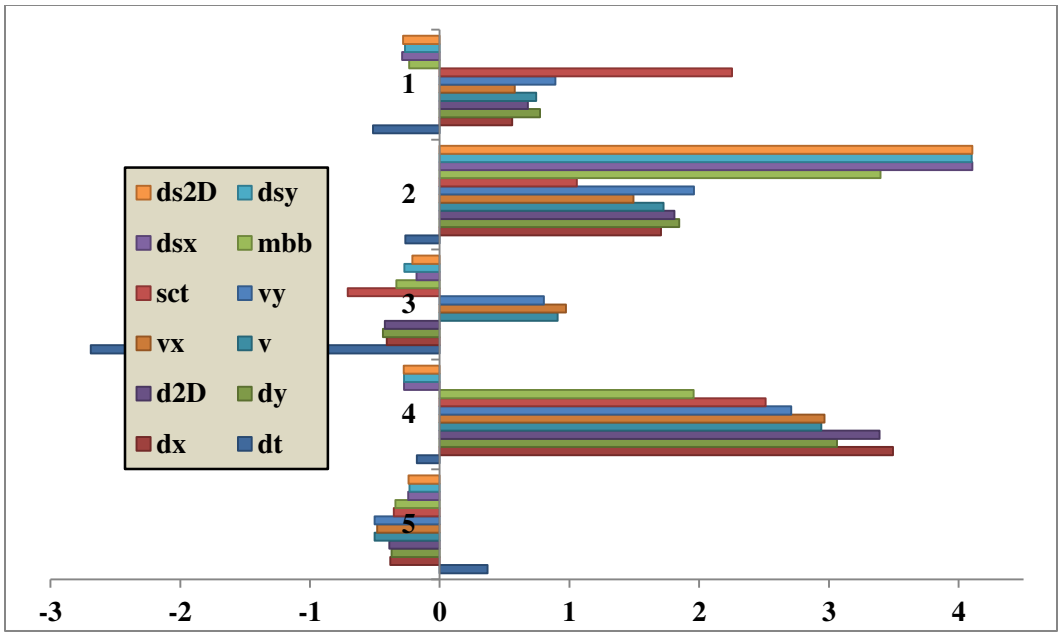

Figure 85. Cluster profile (no partition: *k*=5).

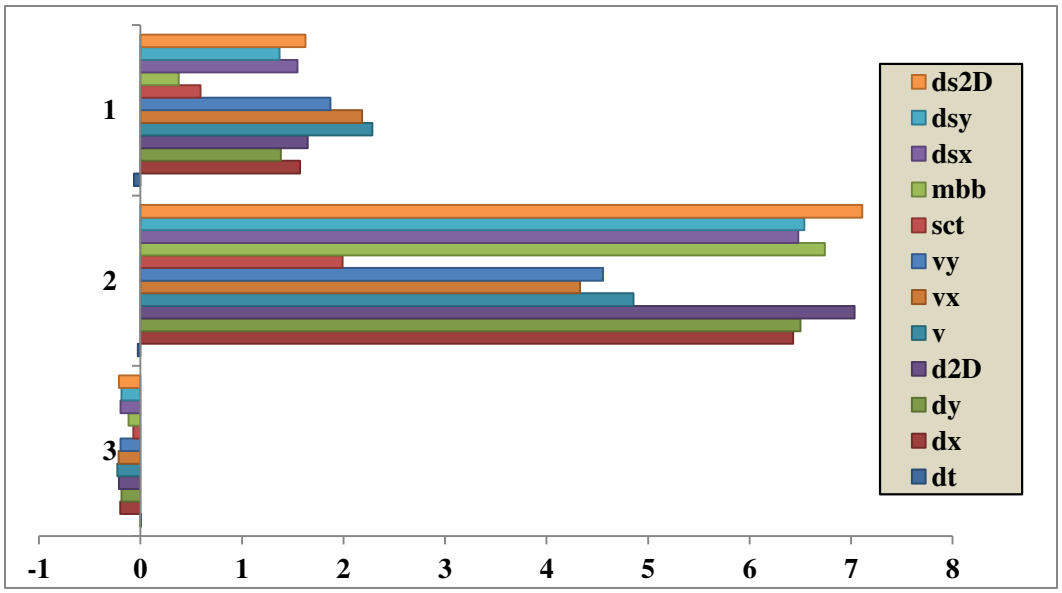

Figure 86. Cluster profile (TRACLUS-MDL: *k*=3).

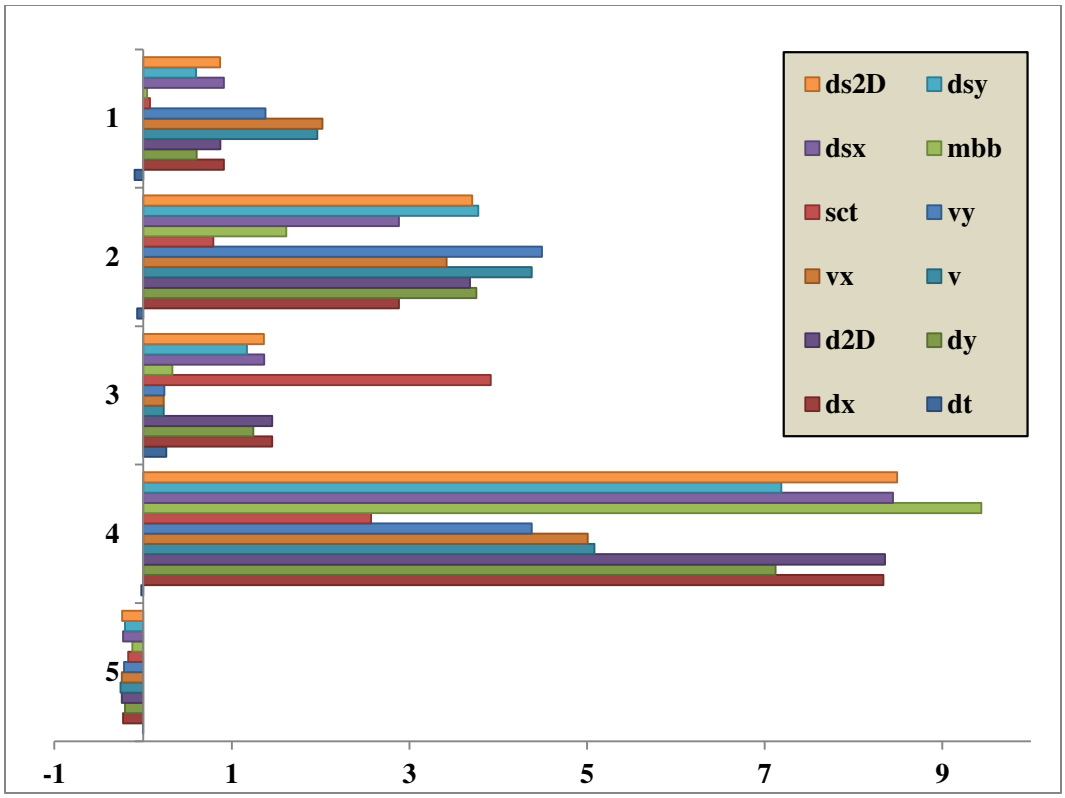

Figure 87. Cluster profile (TRACLUS-MDL: *k*=5).

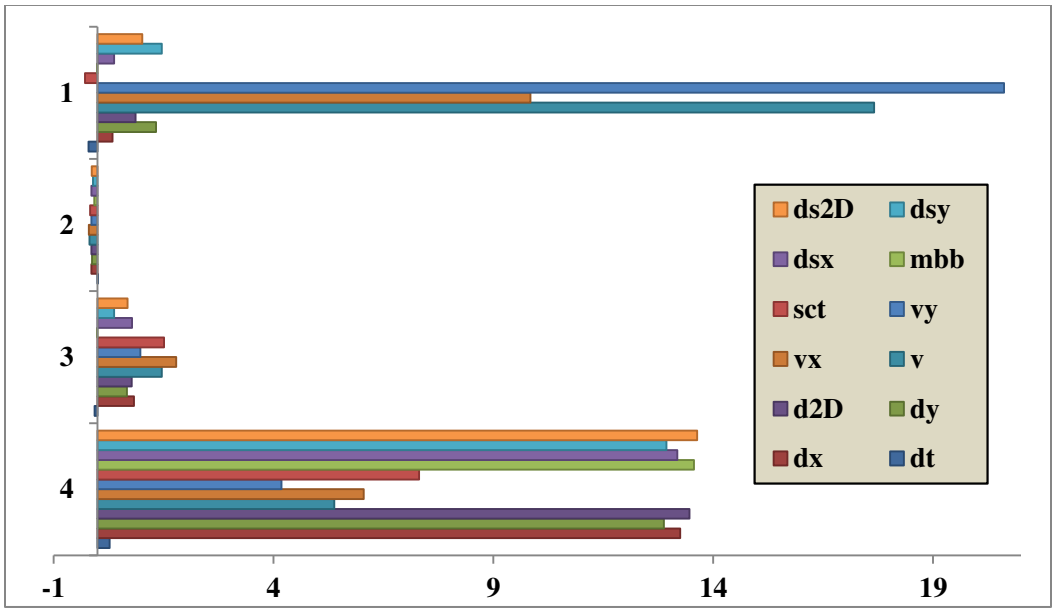

Figure 88. Cluster profile (Distance-Threshold: *k*=4).

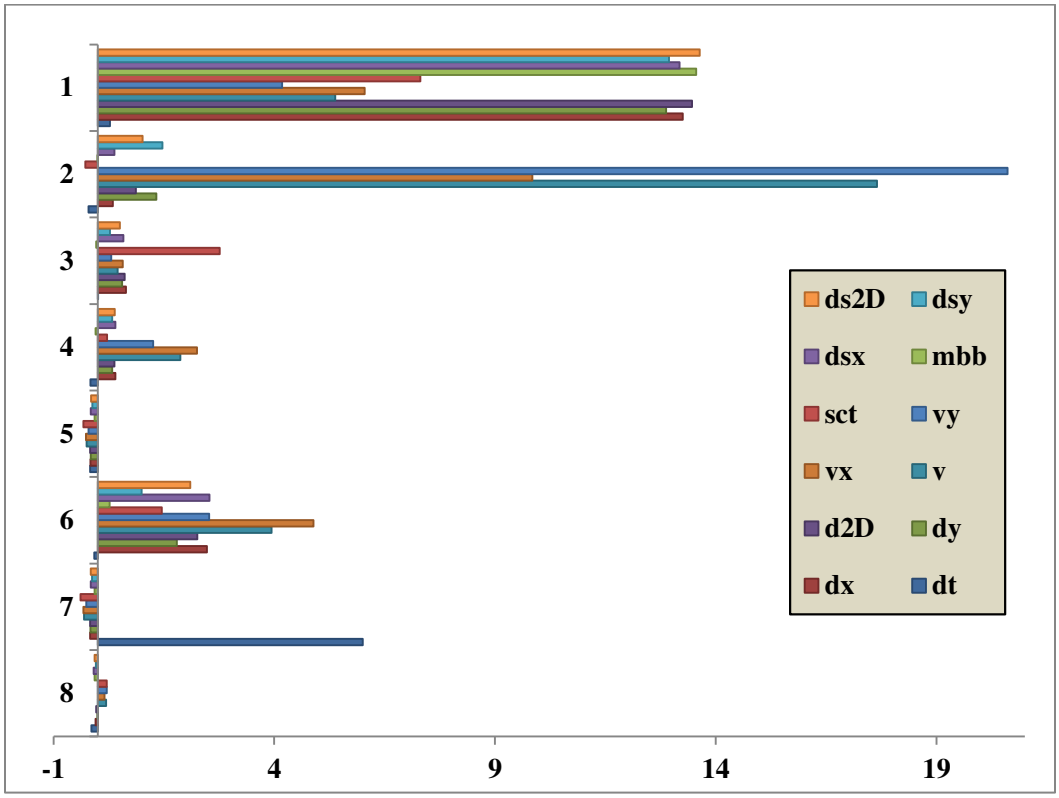

Figure 89. Cluster profile (Distance-Threshold: *k*=8).

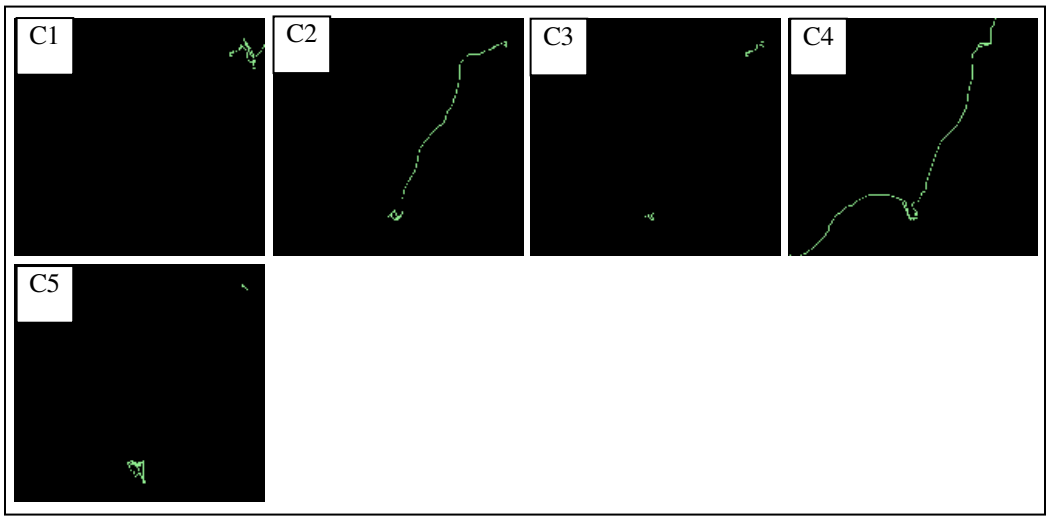

Figure 90. Sub-trajectory clusters (no partition: *k*=5).

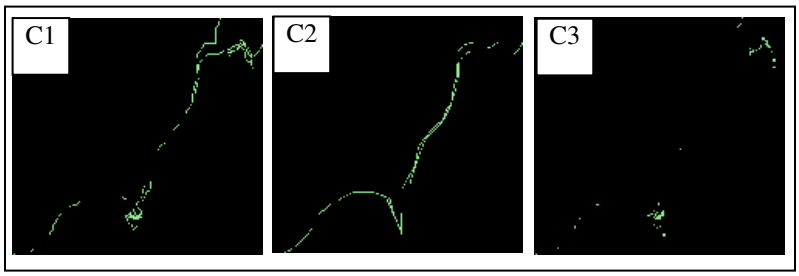

Figure 91. Sub-trajectory clusters (TRACLUS-MDL: *k*=3).

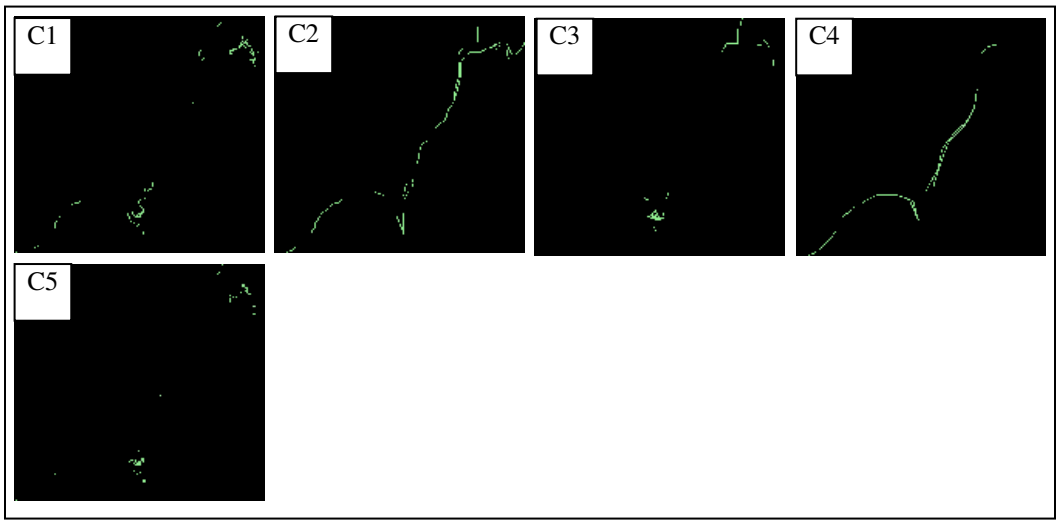

Figure 92. Sub-trajectory clusters (TRACLUS-MDL: *k*=5).

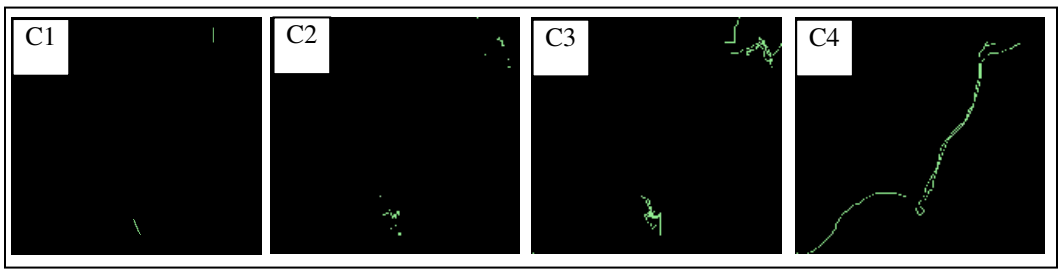

Figure 93. Sub-trajectory clusters (Distance-Threshold: *k*=4).

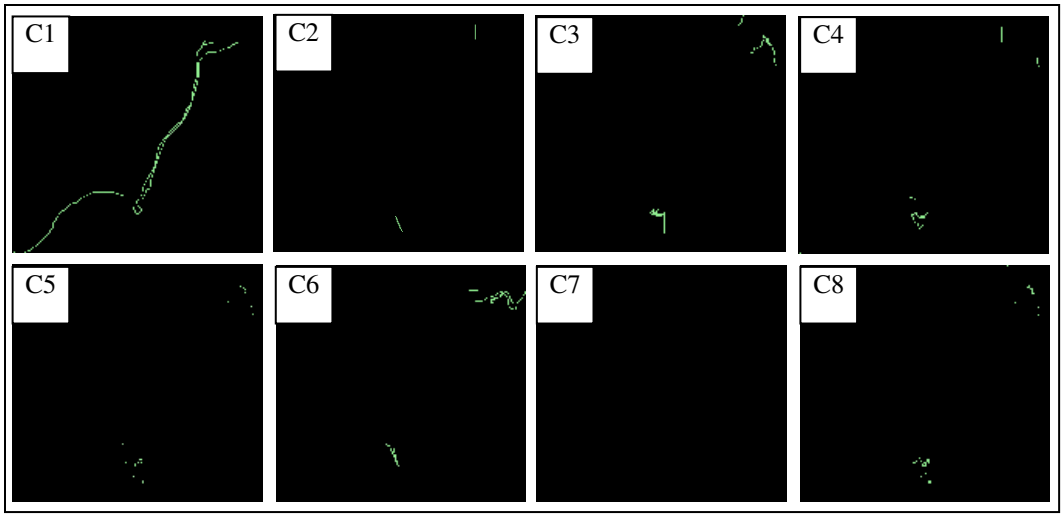

Figure 94. Sub-trajectory clusters (Distance-Threshold: *k*=8).

Figure 95 shows maps of cluster distribution through time for corresponding partitioning methods and selected *k* values. The vertical axis represents each daily trajectory from Day1 to Day36, whereas the horizontal axis is time of day. Each pixel in the images corresponds to a cluster ID at a certain time of day of a trajectory in a 30 seconds interval. In the no partitioning method, each day has only one trajectory; thus it is assigned by only one cluster ID (Top image in Figure 95). 27 trajectories are assigned as Cluster 5 describing staying behavior of working at an office. It matches with the number of "Work" as a label of binary activity in Table 14.

Contrary to the no-partitioning approach, partitioning algorithms extract clusters of local movement behaviors from a daily trajectory. The composition of temporal cluster distribution looks similar in each partitioning algorithm, but more variations appear with large value of *k* (TRACLUS-MDL with *k*=5, Distance-Threshold with *k*=8) that could explain further detail behaviors. For

example, a single staying behavior can be further break-down into a short stay at a grocery store and a long stay at home.

In the images, most trends are drawn by a single sub-trajectory cluster. They are Cluster 3, 5, 2, and 7 for TRACLUS-MDL with k=3, TRACLUS-MDL with k=5, Distance-Threshold with k=4, and Distance-Threshold with k=8 respectively. These clusters represent staying behavior explained by the cluster profiles. According to the travel diary, these behaviors particularly describe working at an office suggesting that the subject of the dataset is a regular daytime office worker. In terms of partition methodologies, these staying behaviors are explained differently. In TRACLUS-MDL with k=3 and k=5, and Distance-Threshold with k=4, the staying behavior is explained by a single cluster (Cluster 3, 5, and 2 respectively), where the cluster profile represents the behavior as low movement (Figure 86 to Figure 88); therefore, the cluster involves not only staying behavior but other low movement behavior such as walking. This result is due to the low number of k suggesting that the optimal k value determined by the gap statistics over-generalized behaviors in this dataset. On the other hand, in Distance-Threshold with k=8, there are three clusters (Cluster 5, 7, and 8) to represent staying behavior and slow movement. Cluster 7 specifically represents long staying behavior with a large duration (Figure 89) that explains majority of working activity. In addition, Cluster 5 represents relatively short staying behavior such as staying at home before turning-off the GPS device, while Cluster 8 represents slow movement such as walking. These three clusters distinguish

detailed staying activities that were not captured by other three partition approaches.

In terms of movement behavior, partitioning approaches can capture commuting behaviors. These behaviors are explained by Cluster 1 in TRACLUS-MDL with k=3, and by Cluster 3 by TRACLUS-MDL with k=5 and two Distance-Threshold approaches. These clusters are regularly found in the morning around 8 a.m. and in the evening with some variation. This reasonably explains a commuting behavior of a typical office worker, who goes to work in the morning at a specified time and leaves his/her office at various time depending on overtime work.

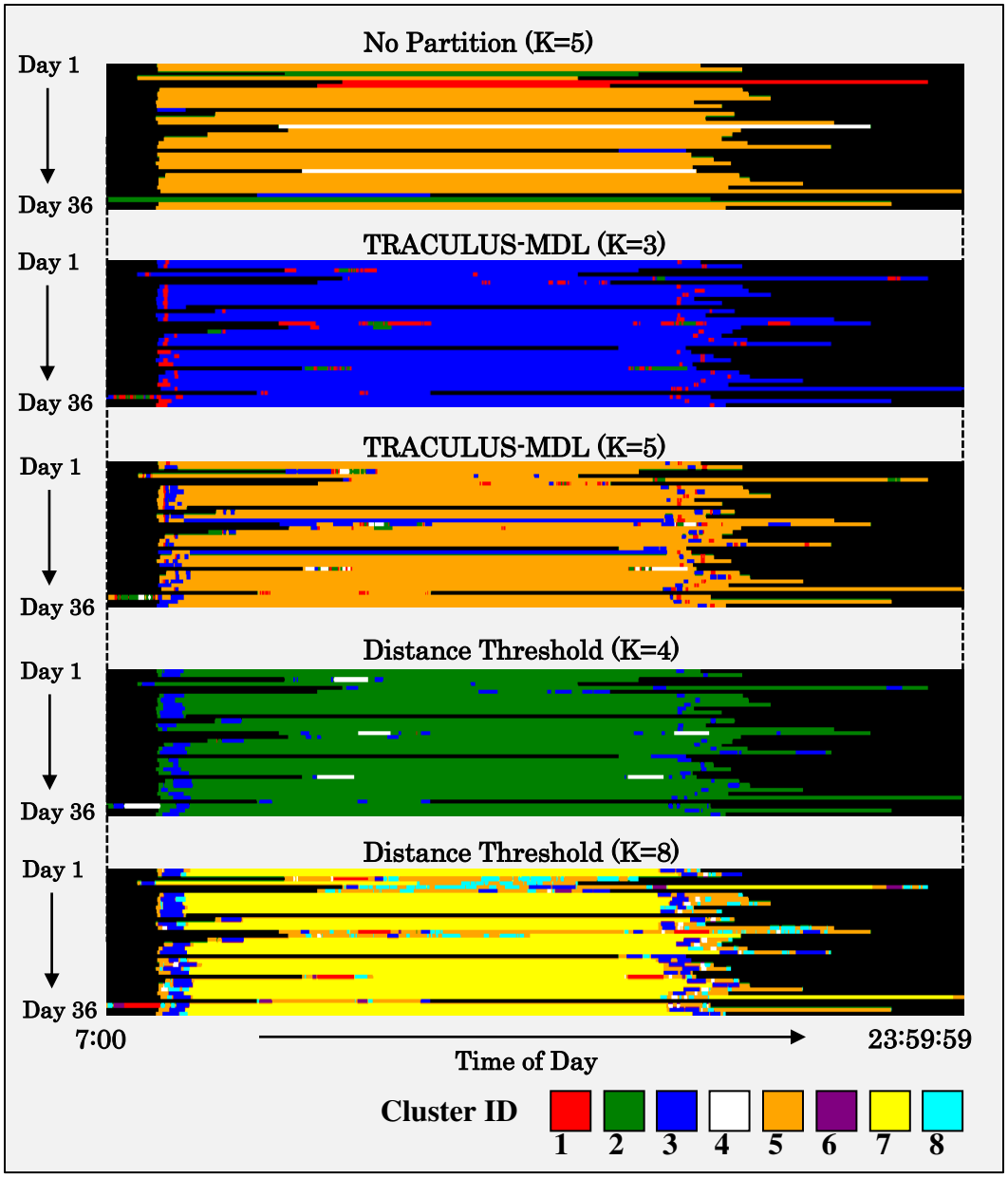

Figure 95 Temporal cluster distribution.

As discussed in 0, J48 was applied to quantitatively confirm how these cluster distributions can explain behavioral contexts of movement. The input dataset for the model was the total time in seconds by each cluster in each day of trajectory and the parameter of confidence factor was set to 0.3. As mentioned in

0, two references of behavioral contexts were used; major activity label and binary activity label. Table 21 and Table 22 present the results of Decision Tree using 10-folds cross validation with the kappa coefficient for the behavioral context recognition of major activity and binary activity respectively. In summary, the recognition of major activity was unsuccessful with low recognition rates and low kappa values for three partitioning algorithms. The result of binary activity recognition was successful with high recognition rates and high kappa values for no-partitioning and Distance-Threshold approaches.

The accuracy and kappa coefficient for recognition of major activity do not show much difference for different partition algorithms or different *k* values. TRACULS-MDL with *k*=5 has the highest recognition rate, 69.4%, while the lowest is 63.9% in no partition, Distance-Threshold with *k*=4, and Distance-Threshold with k =8. In fact, there were only 2 different correctly identified instances between the highest and the lowest recognition. This result means that the explanation power of major activities in the dataset cannot be much improved by different trajectory partitioning approaches. One possible reason is that even though the trajectory data-mining extracted behavioral clusters such as short-long stays, slow-fast movements, and directed-sinuous paths, many different human activities in the real-world can share common activities of such behaviors. For example, "Work&Dining" activity includes behaviors of a work trip in the morning, a long-stay at the office, a trip to the restaurant, a short-stay at the restaurant, and a trip to home, while "Dining" activity on a holiday may be composed of similar behaviors such as a long-stay at home, a trip to the restaurant,

a short-stay at the restaurant, and a trip to home. Because this study only considers the composition of such behaviors explained by trajectory clusters, the recognition of various activities may be limited. Another potential reason is that all three approaches explained trajectories or sub-trajectories based on multiple motion descriptors that are purely based on three-dimensional geometry (x,y,t), and such geometrical explanations cannot fully describe complex behaviors of real-world human activities from the GPS dataset. One potential solution to improve the inference of complex activities is to use other information such as locational information and temporal sequence of trajectory clusters in addition to the composition of trajectory clusters.

On the other hand, the accuracy and kappa coefficient for recognition of binary activity both show high recognition accuracy (except TRACLUS-MDL with k=5). The binary activity categorized a daily trajectory into two simple activities "Work" and "Non-Work". And the daily trajectories labeled with "Work" in this dataset have two common behaviors; a long stay behavior that explain working at an office, and short distance trips that explain commuting behaviors. These behaviors are well extracted by trajectory data-mining particularly for Distance-Threshold with k=8, that results in higher recognition of binary activity. The result also shows no significant difference for recognition of binary activity between no partition and partitioning algorithms. Distance-Threshold had the highest recognition rate 94.4% with k=8, whereas no partition had 91.7% with k=5. This also indicates that partitioning may not largely improve the recognition accuracy in this dataset.

Table 21. Results of decision tree (Main activity).

Partition Algorithm	k	Classification			
		Corr.	Incorr.	Corr. (%)	Kappa
No Partition	5	23	13	63.89	0.32
TRACLUS-MDL	3	24	12	66.67	0.34
	5	25	11	69.44	0.43
Distance-Threshold (D=20)	4	23	13	63.89	0.34
	8	23	13	63.89	0.36

Table 22. Result of decision tree (Binary activity).

Partition Algorithm	k	Classification			
		Corr.	Incorr.	Corr. (%)	Kappa
No Partition	5	33	3	91.67	0.79
TRACLUS-MDL	3	31	5	86.11	0.64
	5	26	10	72.22	0.26
Distance-Threshold (D=20)	4	33	3	91.67	0.77
	8	34	2	94.44	0.85

As a post-analysis, detailed interpretation of trajectory clustering was performed by matching extracted behaviors with actual behaviors, visualizing the decision tree, and visualizing cluster distributions in space and time. Because values of *k* that are too small cannot distinguish various behaviors and because Distance-Threshold with *k*=8 has the highest recognition rate, the analysis was focused on the result of trajectory data-mining using Distance-Threshold partition with *k*=8. First of all, the extracted behaviors of sub-trajectory clusters described by cluster profiles were manually matched with actual behaviors found in the activity diary. Table 23 shows the matching result, and there are three interesting results identified. First, generally extracted local behaviors distinguished different behaviors; however, there are some overlapping behaviors such as “Walk” in both

Cluster 3 and 8. This is reasonable because in reality there can be variations in one term of behavior such as "Walk" that is depended on the situation. In this study, these variations are introduced by describing sub-trajectories based on multiple movement descriptors. Second, behaviors of Cluster 6, describing medium movement, were identified as "Car" and "Subway". Even though GPS signal is missing when the subject is underground, the trajectory data-mining recognizes that "Subway" has similar movement characteristics with "Car" using points immediate before and after underground. Third, the trajectory data-mining captures noisy movements described by Cluster 2, the profile of which shows extreme movements with very high velocity, low *sct*, and short travel distance. These movements are observed when the subject rode a subway or did indoor activities, suggesting that the behaviors are due by measurement errors of such as signal blocks by obstacles and multi-path effects by signal reflection.

Table 23. Behavioral match between clusters and real activities.

Cluster ID	Behavior (cluster profile)	Behavior (activity diary)
1	Fast move & long trip	Express train
2	Extreme move	Signal lost by subway or indoor activity
3	Slow move	Local train, bus, walk
4	Slower move	Local train, subway
5	Stay	Dining, shopping
6	Medium move	Car, subway
7	Long stay	Working at an office
8	Stay & slow move	Shopping, walk, light rail

Secondly, Figure 96 and Figure 97 show tree visualizations of the Decision Tree results for recognition of major activity and binary activity respectively. These show key cluster hierarchy (major activity: 6, 8, 3, 4, binary activity: 7) that can help to describe contexts of behavioral activity. In major activity recognition, the top level of the hierarchy is Cluster 6 that represents movement behaviors by "Car" or "Subway". Whether this behavior was found in a daily trajectory or not classifies if the subject went on a trip. This suggests that the subject does not drive a car or take a subway on a daily basis. The second level of the hierarchy is Cluster 8 that describes stay or slow movements of walking or taking light rail. This means that if this behavior is more than 47 minutes in a day, the behavior of the subject tends to be "Shopping&Dining". This is a reasonable behavior for an office worker living in an urban area because such a person in a work day may not spend much time for shopping due to his/her time budget or does not walk much unless he/she forces him/herself to walk for fitness, for example. The third level of the hierarchy is Cluster 3 describing slow movement. If the behavior is more than 56 minutes in a day, the subject is likely to go out for lunch or dinner after work. The lowest hierarchy is the Cluster 4 that also describes slow movement. Similar to Cluster 3, if the behavior is identified more than 11 minutes in a day, the subject is like to go out for lunch or dinner after work. Otherwise behaviors show typical working day movement. In binary activity recognition, there is only one key cluster that controls the subject's daily behavior. Cluster 7 describes a long staying behavior and if a single day trajectory contains the behavior more than 5 hours, the behavior is recognized as a work day

behavior. These findings are not only useful for behavioral recognition but interesting to characterize person's daily behavioral patterns.

Figure 98 visualizes morning activity patterns in the area of the subject's residence by the Space-Time line density maps (unit: meter × square meter^{-1} × second^{-1}) of corresponding Cluster IDs (output voxel grid size: 200×200 (unit: meter) ×200 (unit: 30 seconds), bandwidth of STKDE: h_1=400 (unit: meter), h_2=400 (unit: 30 seconds)). The images allow overview of the subject's typical morning behavior on a work day. Clusters 5 and 8 in the earlier time explains the subject's behaviors of staying at home after a GPS device was turned on and walking to a train station. Cluster 3 and 4 describe commuting behaviors such as taking a train and walking to the office. Cluster 7 represents working behavior at the office. Cluster 1 (express train) and 2 (signal lost by subway) are irregular patterns of commuting because of low density value.

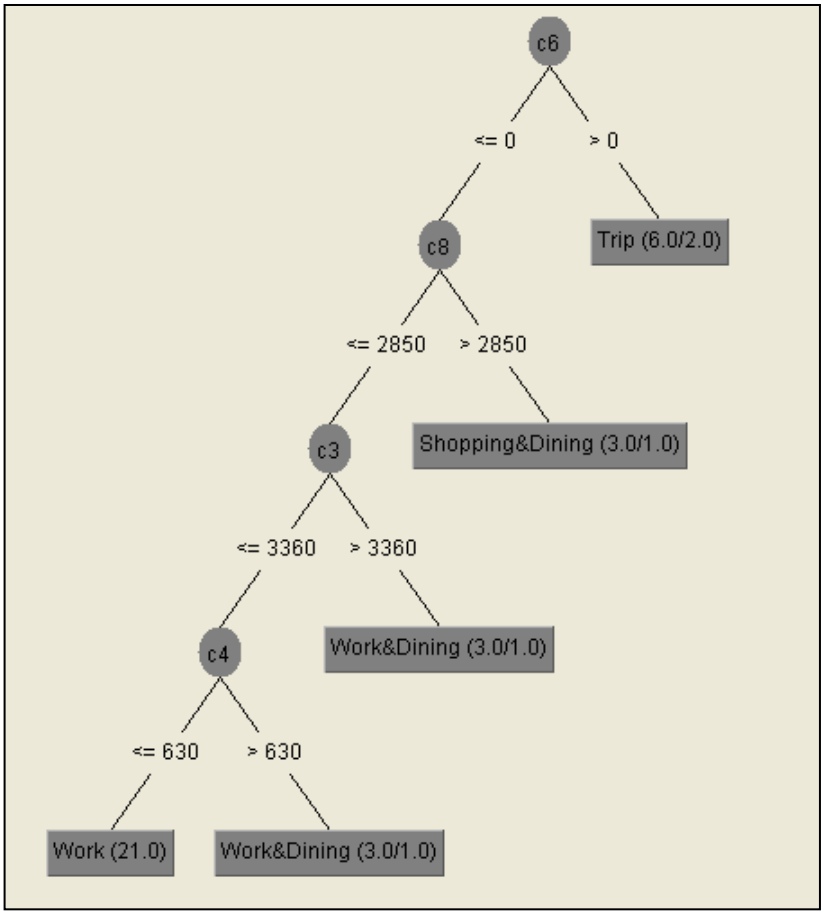

Figure 96. A tree visualization of Decision Tree results (Major activity: Distance-Threshold, *k*=8).

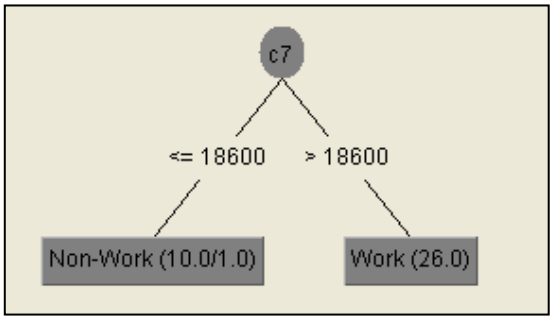

Figure 97. A tree visualization of Decision Tree results (Binary activity: Distance-Threshold, *k*=8).

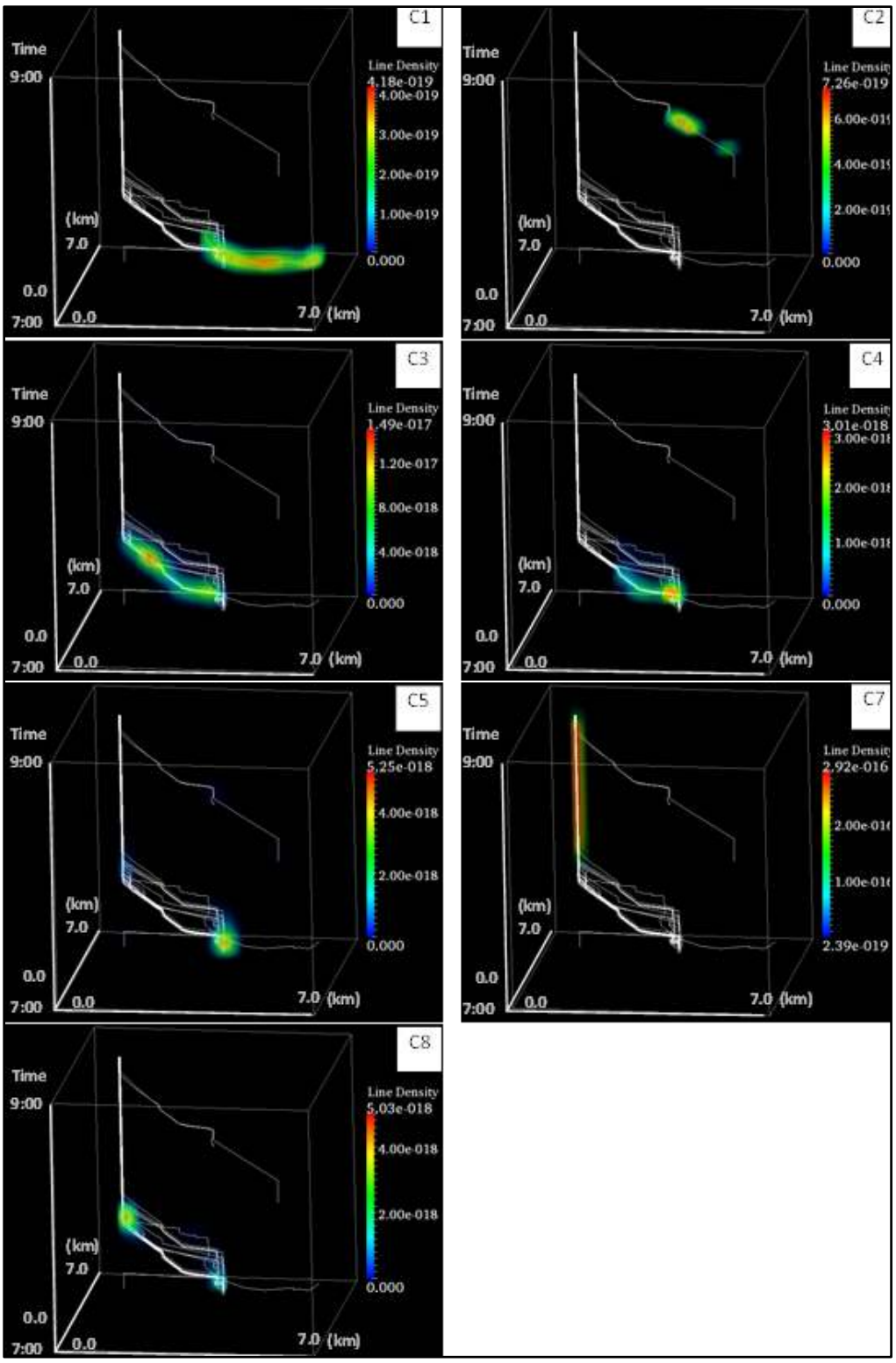

Figure 98. Space-Time line density map in morning activity (Distance-Threshold, k=8).

5.5 Discussion and Conclusions

This research offers insight into three research challenges of trajectory data-mining; 1) how to characterize and generalize massive trajectories to extract interesting patterns; 2) how to explain behavioral contexts of trajectories by those extracted patterns; and 3) how to visualize extracted patterns to overview and compare patterns and trends in space and time. To respond to these challenges, this research developed a novel trajectory data-mining framework and a toolkit. The functionalities of the toolkit include a trajectory data-mining analysis that employs trajectory partitioning and clustering algorithms to extract behavioral patterns of mobile objects and a visual analysis to display extracted patterns and trends in space and time. To examine the capability of the toolkit, two movement datasets were analyzed; 1) mixed movements generated by three different random walk models, BM, CRW, and Lévy flight; and 2) human daily movements in urban space collected by a GPS device.

In summary, the results demonstrated that local behaviors of trajectory were well extracted and can explain the global behavioral context from mixed trajectories of random walkers. Extracted local behaviors in the GPS dataset differentiated real movement activities during a day; however, several behaviors were overlapping in different behavioral clusters. In addition, the explanation power for global behavioral context recognition by local behaviors is not much improved from the recognition by global behaviors. These results indicate that the proposed trajectory data-mining framework performs well on mixed behavioral

datasets that are explicitly defined by mathematical expressions; however, when applied to data from the real-world, the explanation power is relatively limited.

This study also examined the effect of two partitioning algorithms because different partitioning techniques may reveal different behavioral contexts. The TRACLUS-MDL approach partitions a trajectory by finding a sudden geometrical change. Thus, it is particularly useful to extract behaviors when mobile objects show a behavioral change accompanied by their directional change in movements. In the result of the random walk experiment, trajectories of CRW were better partitioned by TRACLUS-MDL because they consist of some directed movements. On the other hand, the Distance-Threshold approach partitions by finding staying behavior along a path, and is useful if behaviors of mobile objects possess staying behaviors. The results showed that the Distance-Threshold approach better partitioned trajectories of Lévy Flight and real GPS datasets.

Two visual analyses to visually confirm the distribution of extracted trajectory clusters are found to be useful. The first analysis is mapping temporal cluster distribution on a 2D bitmap image that allows us to overview how extracted clusters distribute through time for each trajectory in a dataset and help identify similarity and dissimilarity patterns. In the random walk experiment, images of temporal cluster distributions for three partitioning algorithms clearly captured the effect of different approaches. While the no-partitioning approach introduced misclassifications, two partitioning approaches showed that the composition of local behaviors can determine crisp boundaries that distinguished three random walks. In the GPS experiment, the visualization helped to identify

behavioral patterns such as long staying behavior for working at the office and commuting behaviors. The second visualization analysis used maps of STPs and Space-Time line density with STKDE to overview and explore extracted cluster distributions through space and time. These maps are useful in visually confirming patterns and characteristics of extracted behaviors by trajectory data-mining.

There are two major considerations for future work. First of all, more experiments are required to advance the analytical power of the methodology and toolkit; for example, fine-tuning of model parameters particularly concerning spatial and temporal granularity (e.g., resampling frequency, parameters for trajectory partition algorithms, *k* value in k-means clustering, grid size and band width selection for STKDE), variable selection of motion descriptors, methodological exploration with other motion characterization (e.g., incorporating variances in addition to mean values), clustering, and classification techniques, and experiments with other dataset.

Second, despite the agreement between extracted local movement behaviors and actual activities confirmed by the activity diary, the recognition rate of major activity stays around 60 to 70%. One potential reason is that this study didn't account for the temporal sequence information of extracted local behaviors for context recognition, but was purely based on the structural composition of those behaviors. The temporal sequence can be incorporated by using classification techniques that assume the probability theory of Markov process such as Hidden Markov Model and Dynamic Bayesian Networks. In addition,

using multi scale behavioral patterns of trajectory may also improve the analysis. This can be achieved by, for instance, conducting analysis with different sampling frequency. Furthermore, besides trajectory data, other geographical, topological, and personal information can be incorporated into the classification process to improve behavioral recognition accuracy; however it will be a controversial issue between specification and generalization. Over-specification (e.g., adding too much individual-oriented information) may not be appropriate for generalization of collective movement behaviors. Last but not least, treatment of uncertainty in the dataset is another critical issue needed to improve the analysis.

Chapter 6

EVALUATION OF A PEDESTRIAN SIMULATION MODEL BY TRAJECTORY DATA MINING APPROACH

6.1 Overview

During emergency evacuations on streets or in buildings, pedestrian crowds have a chance to encounter secondary disasters, the impact of which causes incidents of serious injuries and fatalities. Potential factors are overcrowding and crushing caused by, for example, street or building structural problems and human stampede behaviors. In fact, such incidents have been reported numerous times every year from around the world (Fruin, 1993; Still, 2000). In order to achieve efficient evacuation of pedestrian crowds from buildings and cities in emergency situations, it is important to analyze the safety of egress design for aspects of structural design as well as pedestrian crowd behaviors.

Pedestrian dynamics and behaviors under emergency situations have been discussed and examined extensively using Agent-Based Models (ABMs), which are a particular type of computational simulation methodology. The simulation framework has a significant advantage for the analysis of egress design. Normally, in order to capture a full understanding of egress design, it requires exposing massive crowds of real people to a specific emergency environment and obtaining empirical data. However, it is hardly feasible due to the high cost in both monetary and security/safety. Therefore, computer-based simulation is a useful alternative tool.

Despite recent success in exploring and developing simulation models for emergency evacuation using ABMs, not many studies have focused on model evaluation to examine how well simulated results represent movement behaviors realistically. In fact, model evaluation has been recognized as one of key research challenges in the field of ABM (Batty & Torrens, 2005; Crooks, Castle, & Batty, 2008).

This study proposes a new analytical framework for evaluating ABMs, not limited to pedestrians but to any mobile objects, by utilizing a trajectory data-mining approach. It extracts detailed spatio-temporal behaviors of mobile objects as a collective movement. The extracted patterns are compared within the framework of time geography by using Space-Time Kernel Density Estimation (STKDE) and three-dimensional map algebra. As a case study, I developed a pedestrian evacuation simulation based on the social force model and generated crowd dynamics on a street corridor with four different scenarios. The evaluation framework is tested to examine simulation dynamics for collective pedestrian movement. The effectiveness of street design is qualitatively and quantitatively investigated.

6.2 Related Works

Modeling pedestrian behavior is an important research topic for many applications, ranging from urban planning (Schelhorn, O'Sullivan, Haklay, & Thurstain-Goodwin, 1999), transportation management (Desyllas, Duxbury, Ward, & Smith, 2003), computational animation (Treuille, Cooper, & Popovic, 2006), to

physical (Helbing & Molnár, 1995), social (Pelechano, Allbeck, & Badler, 2007), behavioral (Timmermans, 2009), psychological (Sakuma, Mukai, & Kuriyama, 2005), medical (Smith, Brown, Yamada, Kowaleski-Jones, Zick, & Fan, 2008), and geographical studies (Torrens, 2011). In particular, modeling pedestrian evacuation dynamics has been extensively studied by scientists and practitioners for safety management for catastrophes (e.g., building fire, street explosion, tornado/hurricane, earthquake, tsunami, terrorist attacks).

Modeling pedestrian evacuation dynamics involves considering many factors including complex human behaviors of physical movement, individual characteristics, spatial environments and configurations, and interactions among pedestrians as well as between pedestrians and the environment at multi-scales in space and time. In addition, model evaluation has been a critical research challenge for a long time; however, not many studies focused on model evaluation to examine how well simulated results represent movement behaviors realistically or in detail, largely due to lack of adequate data.

6.2.1 Pedestrian Movement and Evacuation Behaviors

Pedestrian dynamics consist of complex movement behaviors at multiple scales. For example, macro-scale behaviors of trip planning and activity scheduling (Axhausen & Gärling, 1992; Timmermans & Arentze, 2002), meso-scale behaviors of route choice (Borgers & Timmermans, 1986) and way-finding (Golledge, Klatzky, & Loomis, 1996), and micro-scale behaviors of orientation and locomotion (Montello, 2005). These movement behaviors are also affected by

personal factors such as age, gender, preferences (Bovy & Stern, 1990), past experience (Golledge & Stimson, 1997), the use of mental maps (Kitchin, 1994), space-time constraints (Hägerstrand, 1970), and trip characteristics such as trip purpose (Bovy & Stern, 1990), route structures (e.g., sidewalks, paved, tree, obstacles), a mode of travel (Walton & Sunseri, 2010), and situations along the route (e.g., traffic volume, attractive spots). In addition, non-linear interactions among individuals as well as interactions between individuals and the environment introduce further complexity with feedback, scaling effects, and path dependence.

In the case of evacuation, there are specific factors that affect pedestrian dynamics and behavior under emergency situation. First, the perception of risk is a key factor for an individual's decision to react to a disaster, i.e., to evacuate (Proulx, 2002). An individual's perception of risk often depends on individuals and situations. For example, an individual may not perceive a high sense of risk by a warning system such as alarms if the individual is provided false alarms frequently. A study by Bryan (1995) showed that people do not respond well to non-voice alarms such as bells and sounders. Risk perception also depends on location. For example, in the case of building fire, if individuals are closer to the fire, individuals may perceive a high sense of risk because they can hear noise, smell smoke, and see smoke and fire. However, in other cases such as CBR (Chemical, Biological, Radioactive) disasters, individuals may not perceive risk through their senses due to colorless or odorless materials. Individuals can also perceive risk from the response of others. For example, Latane and Darley (1970)

argued that individuals may downplay the fire cues because some individuals may prefer to evacuate after others around them begin to evacuate.

Second, evacuation response and behavior may also be affected by various characteristics of pedestrians such as physical and psychological conditions, social factors, and knowledge and experiences. For example, some experimental studies and statistical analyses showed age, gender, and disability may have some influence on evacuation timing (Proulx, Latour, McLaurin, Pineau, Hoffman, & Laroche, 1995; Bateman & Edwards, 2002). The effects of panic, which can be defined as a fear-induced flight behavior that is non-rational, non-adaptive, and non-social (Schultz, 1964), have been seen in fire incidents such as the Beverly Hills Supper Club fire (Kentucky State Police, 1977). Emergency egress behavior can also be characterized by social order (Johnston & Johnson, 1988), and roles of individual (e.g., employee, visitor, and leader of a group) can affect how people respond to an emergency evacuation (Bryan, 1982; Proulx, 2002). Social links among members of groups can increase the chance of death because people may delay evacuation or return to the hazardous area in an effort to help one another as the danger of the disaster increased (Feinberg & Johnson, 2001; Cornwell, 2003). Research also showed that individuals' knowledge and experience play an important role for evacuation response and behavior; for example, familiarity with the building and emergency exits, previous experiences in emergencies, and drills (SFPE, 2003). Some studies have shown that, in emergency evacuations, building occupants often exit through the routes that they are familiar with (e.g., the same

route and exit when they entered the building) (Sime & Kimura, 1988; Sime, 1989).

6.2.2 Modeling Pedestrian Dynamics and Evacuation Behaviors

Modeling pedestrian evacuation dynamics is challenging because of the complexity of interrelationship among these multiple factors in determining human movement and evacuation behavior. Many computational models have been developed to simulate pedestrian crowd and evacuation dynamics. There are three approaches commonly used to model crowd and pedestrian dynamics; physics-inspired models, cellular automata (CA), and behavior models.

The first modeling approach, based on physics, was proposed by Henderson (1971), who used an analogy with fluid or gas dynamics to describe how density and velocity of pedestrian flow change overtime, using partial differential equations (Navier-Stokes or Boltzmann-like equations). Hughes (2003) adopted the fluid-based approach to reproduce crowd dynamics, and Treuille, et al. (2006) extended it to model crowds of pedestrians as a continuum flow. Takahashi, et al. (1988) applied the fluid model to simulate building evacuation, in which occupants were treated as a homogenous group with abilities to move with a constant speed, to view the building globally, and to select the most optimal route. Fluid-like crowd behaviors can be observed in the real-word; for example, the footprints of pedestrians in snow look similar to streamlines of fluids, or the emergence of pedestrian streams through standing crowds are analogous to river beds (Helbing, et al. 2002). However, the global behaviors and

homogenous assumptions in continuum models are unrealistic and are not suitable to describe heterogeneous pedestrians who certainly possess local behaviors.

Another physics-inspired approach is based on particle dynamics. Helbing and Molnár (1995) developed the social force model to simulate micro-scale pedestrian motion and crowd dynamics. It described each pedestrian's motion by the summation of forces: a driving force to reach the destination with a desired velocity, repulsive forces to avoid collisions with other pedestrians and obstacles, attractive forces between pedestrians, and fluctuations to introduce stochastic effects. Helbing, et al. (2000) further applied the social force model to simulate panic behavior during pedestrian evacuation. There are a number of advantages; the social force model is mathematically well-described and parameter values of input variables can be measured and calibrated because they have physical meaning. The resulting dynamics (produced by non-linear interactions among pedestrians and their environments with the bottom-up perspective) have the ability to generate self-organizing phenomena (e.g., lane formation, oscillatory flows at bottleneck, stripe formation in intersecting flows, transition to stop-and-go wave, and crowd turbulence (Helbing & Johansson, 2010)) that can be observed in the real-world (Helbing, Buzna, Johansson, & Werner, 2005; Helbing & Johansson, 2007). Nevertheless, the social force model also has drawbacks. In some cases, it generates unrealistic artifacts such as "shake" or "vibrate" behaviors in response to the numerous impinging forces in high-density crowds, which does not correspond to natural human behavior (Pelechano, Allbeck, & Badler, 2008; Torrens, 2011). In addition, although each pedestrian individually

behaves with some stochastic effects, the model scheme applies to pedestrians globally so that the individual behavioral characteristics are not unique. Still (2000, p. 16) argued that "the laws of crowd dynamics have to include the fact that people do not follow the laws of physics; they have a choice in their direction, have no conservation of momentum and can stop and start at will."

Cellular automata (CA) models, an artificial intelligence approach, have been applied for simulating pedestrian dynamics. CA models consist of cells, or grids, that provide the discrete confines of individual automata. Cells own a finite set of states that is used to describe pedestrian attributes such as individual/group occupancy status and their characteristics, and environmental attributes such as room, floor, and obstacles. At each discrete simulation time step, the states of each cell evolve according to well defined uniform transition rules that are locally applied (i.e., the cell itself and its neighbors). An example of CA evacuation models is EGRESS (AEA Technology, 2002), which is based on hexagonal grids and has been applied to simulate evacuation under a variety of circumstances such as fire and smoke. There are several limitations of CA in representing spatial dynamics of pedestrians; for example, because pedestrians are placed on grids and their movement is controlled by probabilistic choices during evacuation, they can unrealistically move in all directions without considering social behavior, personal space, initial speed and movement (Muhdi, 2006). In addition, traditional rectangular grids CA models produce chess-like pedestrian movements. Furthermore, simulation behaviors rely on the choice between two updating schemes, synchronous and a synchronous (Torrens & Benenson, 2005). In the

synchronous updating system, all cells are assumed to change simultaneously, which produces conflicts as in the case of two pedestrians trying to move to the same grid. In the asynchronous updating system, cells change in turn, with each observing a geographic reality left by the previous automata so that the conflict in the synchronous updating system is resolved. The order of updating can be selected as randomly or sequentially in order of some characteristics; however, the updating method is critical as it may influence simulation results (Torrens & Benenson, 2005).

One criticism of the above-mentioned crowd models is that they treat pedestrians as having the same behavior and ignore individual heterogeneous characteristics (e.g., personality, preference, emotion, relationship) (Braun, Musse, de Oliveira, & Bodmann, 2003). In response, a number of behavioral crowd simulations have been developed often using Multi-Agent Systems (MAS). The elemental component of the system is autonomous agents. Franklin and Graesser (1996, p. 25) offer an intuitive description of agents: "An autonomous agent (1) is a system situated within and a part of an environment; (2) that senses that environment and acts on it, over time; (3) in pursuit of its own agenda, and (4) so as to effect what it senses in the future."

For example, Reynolds (1987) developed a crowd model based on flocking and steering behaviors. The flocking mechanism consists of three behaviors; collision avoidance, velocity matching, and flock centering. Under these behaviors, autonomous agents avoid collisions with nearby flockmates, attempt to match velocity with nearby flockmates, and attempt to stay close to

nearby flockmates. Reynolds (1999) also presented a model of steering behaviors by three hierarchies of motion behaviors: action selection, steering, and locomotion. Action selection involves strategy, goals, and planning for autonomous agents' motion behaviors. Steering behaviors model navigation process for an autonomous agent. Specifically, Reynolds (1999) implemented six steering behaviors, including seek, pursue, wander, follow paths, avoid obstacles and follow flows. Locomotion represent agents' embodiment, which converts signals from the steering layer into motion of the character's body (Reynolds, 1999).

Other approaches consider psychological, physiological, and sociological aspects of crowd behaviors (e.g., Egges, et al. 2003; Pan, et al. 2006; Pelechano, et al. 2007). Pelechano, et al. (2007) developed a MAS called HiDAC (High-Density Autonomous Crowds), which incorporated physiological and psychological behavioral factors on top of the social force model. In the model, agent behaviors are computed at two levels; 1) high-level behaviors including navigation, learning, communication between agents, and decision-making; 2) low-level behaviors describing perception and a set of reactive behaviors for collision avoidance, detection, and response to move within a bounded space (Pelechano, Allbeck, & Badler, 2007). The model successfully generated realistic crowd dynamics including bi-directional flows, fire evacuation with panic situations, and high-density crowds under calm conditions.

Durupinar, et al. (2011) extended the HiDAC by integrating a personality model, Five Factor Model (FFM), which is a popular approach in psychology.

FFM describes personal characteristics based on five factors, OCEAN; Openness, Conscientiousness, Extraversion, Agreeableness, and Neuroticism (Digman, 1990). To computationally implement OCEAN, Durupinar, et al. (2011) converted the five factors into 13 low-level parameters; leadership, trained/not trained, communication, panic, impatience, pushing, right preference, avoidance/personal space, waiting radius, waiting time, exploring environment, walking speed, and gesturing. Results of crowd dynamics were evaluated by creating 15 animations presenting OCEAN behaviors and finding the correspondence between the animation and users' perception of the animated behavior using a questionnaire. The result of visual-based evaluation indicated that the model explains five factors well (except "Conscientiousness") because of high correlations between model dynamics and users' perceptions. These mixed scheme behavioral models have significant advantages to generate complex and realistic crowd behaviors.

6.2.3 Model Evaluation

Despite the fact that ABMs have flourished in the field of crowd studies, model evaluation is a long standing issue and still one of the most difficult tasks of doing research in crowd simulation. Model evaluation involves examining how well simulated results represent real-world dynamics; however, it is difficult to compare a simulation to the real-world and its real characteristics because reality involves very complex behaviors. The fundamental question is which aspects of crowd dynamics from simulation are to be compared with real-world dynamics.

Furthermore, the lack of reliable and sufficient data from the real-world is a major barrier especially when evaluating pedestrian evacuation models.

Model evaluation, specifically calibration and validation, of microscopic pedestrian models can be performed by comparing aggregate model outcomes, predicted macroscopic relations, or emerging spatio-temporal patterns with macroscopic empirical data (if available) or expert opinion (Hoogendoorn, Daamen, & Landman, 2005).

Comparison using aggregated variables such as flows, speeds, densities, and overall evacuation time is the simplest approach to capture global behaviors between a model outcome and real dynamics. Many evacuation models, particularly commercial ones, have been validated by comparing overall evacuation time with the estimated overall evacuation time from, for example, evacuation drills. This is useful because overall evacuation time is the primary interest for practical applications in evacuation management; however, aggregated variables cannot explain detailed spatio-temporal process of crowd dynamics.

Another approach is to use a fundamental diagram that explains the relation between density and flow or velocity (Seyfried, Steffen, Klingsch, & Boltes, 2005); this has been used for the evaluation of pedestrian models (Helbing & Molnár, 1995; Hoogendoorn & Bovy, 2000; Keßel, Klüpfel, Wahle, & Schreckenberg, 2001). Studies showed that even though the velocity-density relation differs depending on pedestrian facilities such as corridors, stairs, or halls, the fundamental diagram is associated with every qualitative self-organization phenomena like lane formation or occurrence of congestions (Seyfried, et al.,

2010). However, there is no general consensus about specifications in different experimental studies, guidelines, and handbooks, even for the most relevant characteristics such as maximal flow values, corresponding density, or the density at which flow is expected to become zero due to overcrowding (Seyfried, et al., 2010).

As an example of using flow and density for validation, Penn and Turner (2002) compared the flow rate between simulation and the real-world. Penn and Turner (2002) developed a pedestrian simulation model for a department store by applying Visibility Graph Analysis (VGA), which is derived from the concept of space syntax (Hillier & Hanson, 1984), to model agents' movement. They evaluated model performance by comparing the flow rate between observation data and simulation at 49 gates, where the unit of flow rate is the number of pedestrian per hour at a gate. The result showed that the correlation between agent movement and observed shopper movement had a positive correlation (r^2=.56, n=49, p<.001). Berrow, et al. (2005) used pedestrian flow and density distributions at several congested areas (e.g., entrance at stadium, boarding area at a metro station) over time to compare simulation outcomes from the Legion model (Still, 2000) to observations. The results indicated that no general pattern for the flow-density relationship exists so that strong context-dependencies need to be factored into any attempt to model crowd patterns (Berrow, Beecham, Quaglia, Kagarlis, & Gerodimos, 2005).

Although aggregated variables of simulation outcomes and the fundamental diagram can capture important characteristics of crowd dynamics,

these consider only basic properties about movement. Crowd dynamics are far more complex due to the interaction between pedestrians and their surrounding environments and situations, collective behaviors, individual decision-making process, psychological elements, individuals' knowledge and experience, communications and/or space and temporal scaling effects.

To incorporate spatial scaling effects, Torrens (2011) applied fractal dimension analysis to compare trajectories between real-world traces and simulated trajectories. While trajectories from real-world were collected by GPS, simulated trajectories were generated by a Geographic Automata model, the functionality of which has rich movement behaviors at three hierarchical scales (macro-, meso-, and micro-scale). Fractal dimension analysis specifically compared the movement behavior in different spatial scales in terms of sinuosity and scale-invariant effect.

Schadschneider, et al. (2008) listed self-organized collective behaviors, which can be observed in pedestrian crowd and evacuation dynamics; for example, jamming, clogging, and zipper effect at bottlenecks; stop-and-go waves in high density crowds; lane formation in counterflow; oscillations in counterflow at bottleneck (e.g., doors); roundabout behavior at intersections; and panic (i.e., non-adaptive behavior such as selfish, social, irrational behavior) in emergency situations. These collective behaviors should be concerned when evaluating crowd models; however, there is no sophisticated method to quantify such behaviors for model evaluation.

Visualization-based comparison is also a common approach to compare complex movement behaviors. For example, as mentioned in the previous section (0), Durupinar, et al. (2011) compared animations of crowd simulation and user's perception of the animations to evaluate complex behavioral crowd dynamics.

Another major concern about model evaluation, particularly for evacuation models, is the lack of reliable and sufficient data from the real-world. Even though with advances in camera technology, computer vision techniques for automatic pedestrian detection, and camera devices and location-aware sensors ubiquitously distributed in urbanized areas, data about pedestrian evacuation dynamics from a real disaster are rarely available. Instead, data from experiments or evacuation drills are typically used for model evaluation. However, such empirical data usually do not fully reflect real evacuation dynamics due to practical, financial, and ethical constraints.

6.2.4 Research Objectives

As discussed above, evaluation of crowd models has not been explored sufficiently. Particularly, model validation is a difficult task when systems in the real-world as well as these generated by ABM exhibit complex behaviors, such as feedback, path-dependence, phase shift, non-linearity, emergence, adaptation, and self-organization. A research challenge is which aspects of model behavior ought to be compared with empirical data. Complex behaviors cannot be simply examined by looking at global statistics; it is necessary to consider spatio-temporal process and behaviors across various scales. Developing an analytical

framework for model comparison to empirical data is also useful to compare simulation outcomes to what-if scenarios.

This study proposes a new analytical framework for evaluating ABMs of pedestrian (or any mobile objects). The developed framework is specifically focused on model validation in order to extract detailed spatio-temporal behaviors of mobile objects as a collective movement. It utilizes a trajectory data-mining technique that uses trajectories of mobile objects from real-world and ABMs as input datasets, partitions the trajectories into sub-trajectories, and identifies behavioral clusters based on their motion characteristics. The extracted patterns will be compared and visualized under the concept of time geography using STKDE to exploratory investigate spatio-temporal patterns and trends. Furthermore, three-dimensional map algebra is employed to compare similarity/dissimilarity in behavioral patterns between real and model, between different models, or between different scenarios. As a case study, an ABM of pedestrian evacuation based on the social force model is developed. It generates crowd evacuation dynamics on a street corridor with four different scenarios. Then the proposed framework is applied to quantitatively and qualitatively evaluate the dynamics in different scenarios in order to investigate the evacuation effectiveness of street design.

6.3 Methodology

To examine the proposed trajectory data-mining scheme for evaluating crowd models, I developed a pedestrian evacuation simulation based on the social force

model. The rational for selecting the social force model is that its capability to generate complex dynamics from non-linear interactions of pedestrians and its tractability by well-understood mathematical models. In addition, homogeneous behavior of pedestrian movement can be appropriate for explain a certain evacuation dynamics because the movement of a crowd is more straightforward in the case of an emergency (i.e., go to the exit) than in the general case such as wandering at a shopping mall.

6.3.1 Pedestrian Evacuation Simulation based on Social Force Model

The social force model, a physics-based model for pedestrian dynamics was developed by Helbing and Molnár (1995); it is closely related to gas-kinetic and fluid dynamics. The model is based on assumptions that a mixture of socio-psychological and physical forces influence behavior in a crowd (Helbing, Farkas, & Vicsek, 2000): Each of N pedestrians *i* of mass m_i likes to move with certain desired speed v_i^0 in a certain direction e_i^0, and therefore tends to correspondingly adapt his/her actual velocity v_i with a certain characteristic time τ_i and random behavioral variations $\xi_i(t)$ (Helbing, Farkas, & Vicsek, 2000).

$$m_i \frac{dv_i}{dt} = m_i \frac{v_i^0(t)e_i^0(t) - v_i(t)}{\tau_i} + \xi_i(t)$$

The above equation represents Newton's second law of motion. This specifies that a force that generates pedestrians' movement depends on a mass of pedestrian *i* multiplied with an acceleration (or change in velocity in time) of pedestrian *i*.

Simultaneously, the agent tries to keep a velocity-dependent distance from other pedestrians *j* and walls *w*, and the equation can be rewritten as follows (Helbing, Farkas, & Vicsek, 2000).

$$m_i \frac{dv_i}{dt} = m_i \frac{v_i^0(t)e_i^0(t) - v_i(t)}{\tau_i} + \sum_{f(\neq \mathrm{i})} f_{ij} + \sum_w f_{iw} + \xi_i(t)$$

where, $\sum_{j(\neq i)} f_{ij}$ is a repulsive interaction force describing the psychological tendency of two pedestrians *i* and *j* to stay away from each other and $\sum_w f_{iw}$ is an interaction force with a wall. $\sum_{j(\neq i)} f_{ij}$ and $\sum_w f_{iw}$ are further broken down as follows (Helbing, Farkas, & Vicsek, 2000).

$$f_{ij} = \{A_i exp[(r_{ij} - d_{ij})/B_i]\}n_{ij}$$

where, $r_i(t)$ is the change of position by velocity $v_i(t)=dr_i/d_t$, A_i and B_i are constants, $\{A_i exp[(r_{ij} - d_{ij})/B_i]\}n_{ij}$ is a repulsive interaction force, d_{ij} is the distance between the pedestrians' centers of mass, and n_{ij} is the normalized vector pointing from pedestrian *j* to *i*.

$$f_{iw} = \{A_i exp[(r_{iw} - d_{iw})/B_i]\}n_{iw}$$

where, d_{iw} is the distance to wall *W*, and n_{iw} is the direction perpendicular to it.

Pedestrians in this basic form of the social force model walk unidirectionally, i.e., each pedestrian agent travels between its origin and its destination. To overcome the deficiency, the idea of multiple waypoints is implemented. In the algorithm, each pedestrian (*i*) owns a sequenced list of waypoints and walks toward the first waypoint in the list. When he reaches the waypoint within a certain buffer zone described by a two-dimensional vector *bZ*(*bx*, *by*), the waypoint is removed from the list and walks toward the first

waypoint in the new list until reaching the final destination. These multiple waypoints can be generated for each pedestrian by various path-planning algorithms such as a hill-climbing algorithm, Dijkstra's algorithm, and A* search.

6.3.2 Trajectory Data-Mining for Evaluating Pedestrian Dynamics in Agent-Based Model

ABM simulations of mobile objects generate massive trajectory datasets. In order to evaluate ABMs, this study proposes a new evaluation framework specifically focusing on model validation, i.e., comparing model structure and outcomes to measure goodness-of-fit. The proposed framework is based on analytical examination of movement behaviors of agents in space and time by utilizing trajectory data-mining and time geography visualization. A schematic overview of the framework is illustrated in Figure 99.

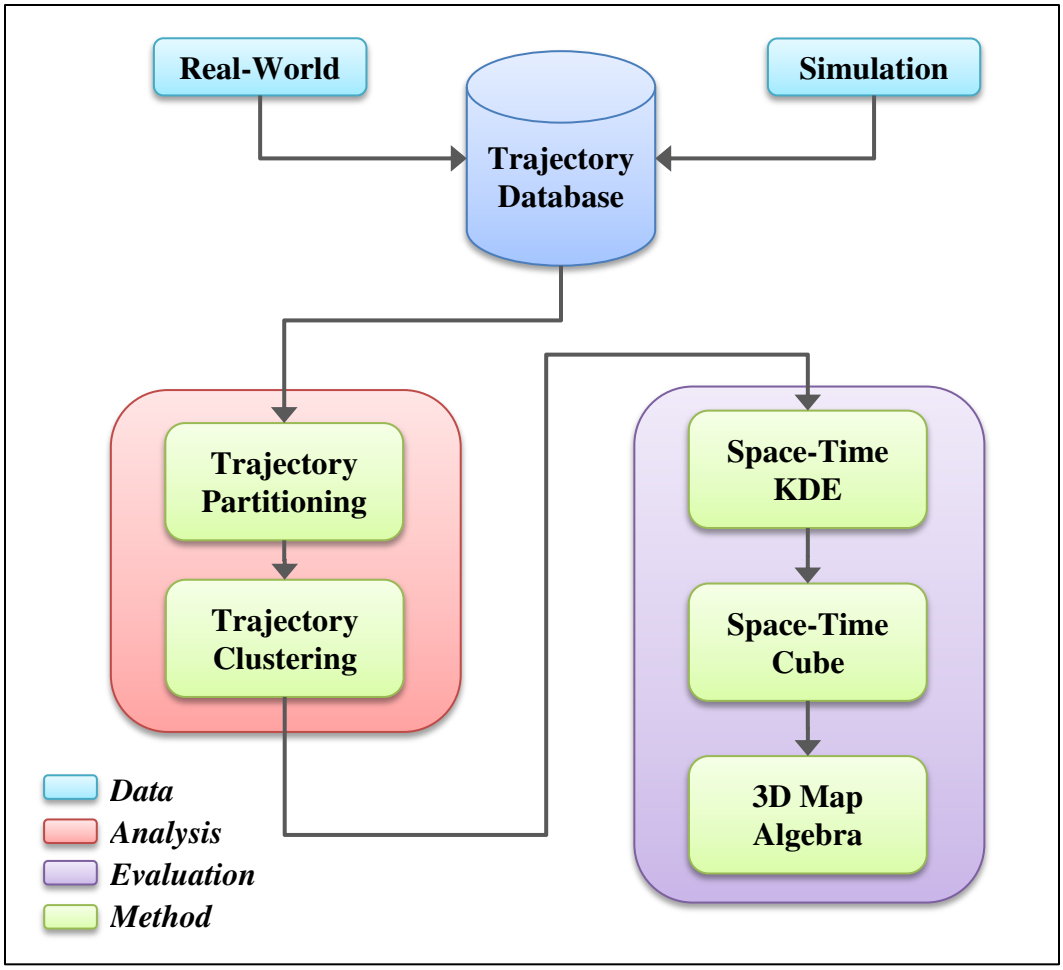

Figure 99. Schematic overview of evaluation procedures for ABMs of mobile objects.

To compare movement behaviors in ABMs, the first process is to merge trajectory datasets into one table in a database. When validating an ABM against the real-world, trajectory datasets should be collected from real-world and an ABM should generate trajectories with the same spatial and temporal units as the real-world dataset. Correspondence of spatial and temporal units is also required when comparing between different ABMs or different simulation scenarios.

The second process involves extracting movement behavioral clusters locally in space and time from the merged trajectory dataset. It consists of two procedures, trajectory partitioning and trajectory clustering. Trajectory partitioning partitions a single trajectory in the merged trajectory dataset into a set of sub-trajectories, while trajectory clustering uses the sub-trajectory dataset and groups them with similar motion characteristics (see 5.3).

The rationale for the trajectory partitioning and clustering approach is to explain movement behaviors of mobile objects in detail in space and time. In the real-world, a trajectory of a mobile object may have a long and complicated path so that it potentially holds various movement behaviors in space and time. For example, a person's daily trip may be composed of multiple transportation modes such as walk, run, and vehicle, while a daily animal path may consist of wandering movement for foraging activity and rapid escape behavior when it is chased by a predator. As a micro-scale movement behavior, a trajectory of a pedestrian on a street corridor may be composed of walking at desired speed, avoiding other pedestrians and obstacles, and queuing until congestion is cleared. Trajectories accompanied with these various behaviors may also depend on a specific time (e.g., time of day, day of week, season of year), space (e.g., street design, infrastructures, landscapes), and situation (e.g., crowd density, panic). When an algorithm clusters trajectories as a whole, it cannot detect similar portions of trajectories because even though some portions of trajectories show a common behavior, the whole trajectories might not (Lee, Han, & Whang, 2007). Therefore, it is important for a trajectory clustering algorithm to have the ability

to detect and group similar portions of trajectories in order to identify local movement behaviors. The second component of my work (Chapter 5) also proved that partitioning approaches better explain the behavioral contexts underlying trajectory datasets. Furthermore, detection of local movement behaviors is particularly useful to determine if collective movement behaviors in trajectory datasets exist. In this study, between two trajectory partitioning algorithms proposed in Chapter 5, I chose to use the Distance-Threshold approach because the focus of this research is to compare behavioral patterns of sub-trajectories rather than geometrical patterns.

Another aspect of the work is the evaluation of ABMs of mobile objects based on extracted behavioral clusters of sub-trajectories. I will introduce two approaches; 1) comparing temporal distribution of sub-trajectory clusters by visual and statistical analyses, and 2) comparing spatio-temporal distributions of sub-trajectory clusters by visual analysis using STKDE and three-dimensional map algebra. The comparison of two types of cluster distributions allow us to evaluate ABMs of mobile objects through identifying similarity/dissimilarity in behavioral patterns between real and modeled scenarios, between different models, or between different scenarios.

To evaluate ABMs of mobile objects, I propose to compare the temporal and spatio-temporal distribution of sub-trajectory clustering. I apply various visualization techniques because the human visual system is extremely effective at recognizing patterns, trends, and anomalies (Miller & Han, 2009).

In the first approach, temporal distributions of sub-trajectory clusters in different trajectory datasets (e.g., real-world, ABMs) are compared by visual and statistical analyses. For visual analysis, the temporal cluster distribution can be mapped on a 2D bitmap image, where an x axis represents time, an y axis represents each pedestrian ID, and each pixel is colored by Cluster ID. For statistical analysis, correlation is employed to find the relationship among sub-trajectory clusters. The combination of visual and statistical analysis can answer questions regarding the behavioral patterns and process of collective movement in different ABMs. For example, what is a cluster and why is a particular sub-trajectory cluster identified in one simulation but not in others?; and what is the cause and effect relationship between/among sub-trajectory clusters in relation to movement behavior through time? Answering these questions allows us to evaluate behavioral components in ABMs (and perhaps in reality).

Another thing that I will show is how to evaluate ABMs by examining how trajectory clusters of mobile agents are distributed through space and time. To accomplish this, I employed STKDE (see details in Chapter 4) and three-dimensional map algebra. Using STKDE, spatio-temporal cluster distributions can be mapped in a 3D space-time cube where the x-y axis represents geographical positions and z axis to represent time. STKDE estimates a point density distribution of sub-trajectory clusters in a space-time cube, and a volume rendering technique allows visual analysis to find similarity and dissimilarity of distribution patterns in different ABMs.

In order to specifically focus on the visualization of dissimilarity, I employed the idea of three-dimensional map algebra. Map algebra, first introduced by Tomlin and Berry (1979), is a two-dimensional raster-based analytical language. Map algebra operators are generally the same operators found in scientific calculations such as arithmetic, relational, boolean, logical, and combinational. In addition, Tomlin (1990) defined several high-order operations, which are typically organized into three major functions; local, focal, and zonal. Local functions create an output grid where every single output cell value is computed from the values of the same location in one or more input grids (i.e., on a per-cell basis). Focal functions compute values in the output grid that are determined by the center cell and its specified neighbors in input grid(s). Zonal functions create an output grid where the output value for each location is a function of the values from an input grid (the value layer) that are associated with that location's zone on a reference grid (the zone layer).

Although map algebra operations are relatively simple, the combination of many operations makes map algebra a rather powerful tool to perform complex tasks, and thus it has been incorporated in commercial GIS and remote sensing software. Furthermore, a number of extensions to map algebra have been proposed; for example, Takeyama & Couclelis (1997) integrated map algebra and cellular automata to incorporate spatial dynamics (GeoAlgebra), Ledoux and Gold (2006) applied the Voronoi diagram instead of a regular tessellation, and Mennis, Viger, and Tomlin (2005) extended two-dimensional map algebra to three-

dimensional cubic map algebra to handle spatio-temporal datasets within the framework of time geography.

This study simply utilizes local and subtraction operators to compare between two space-time cubes in which spatio-temporal distributions of sub-trajectory cluster are estimated by STKDE (Figure 100). Because the process of trajectory data-mining requires merging trajectory datasets from different simulations and space-time cubes of sub-trajectory cluster density distribution are derived from the merged dataset, the spatial extent, resolution, and orientation of voxel grids in space-time cubes for each cluster distribution are the same. Therefore, a three-dimensional map algebra operation can be directly applied without any further resampling procedure, which typically degrades information. Space-time cube visualization based on the outcome of three-dimensional map algebra operation allows further investigating spatio-temporal dissimilarity in movement behavior described by sub-trajectory cluster distributions between two simulations.

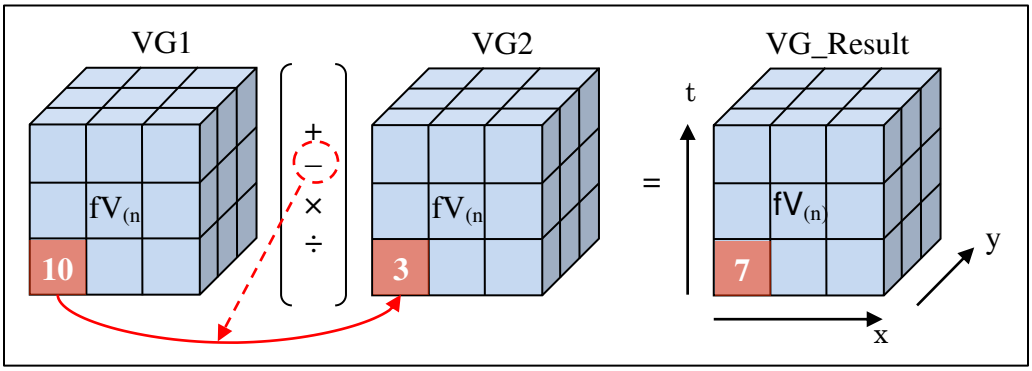

Figure 100. Three-dimensional map algebra using local function and arithmetic operator of subtraction.

6.4 Results

6.4.1 Simulation Scenarios and Dataset

In this study, pedestrian evacuation dynamics on a four-way intersection are simulated using the social force model. Intersections are used because mutual obstructions are practically unavoidable, and therefore street design and evacuation management are important for, for example, organizers of mass events (Helbing, Johansson, & Lämmer, 2007). Simulations are useful to examine the flow performance in different designs of a four-way intersection. Johansson (2008), for instance, compared the flow performance of pedestrians between conventional and improved designs of intersections. In the study, the conventional design is a simple four-way intersection with right angle corners, whereas in the improved design, three feature (railings in corridors, a pillar in the middle, and rounded corners) were added to encourage circular traffic, i.e., roundabout effects (Helbing & Molnár, 1997). The result showed that the flow rate became twice as high by improving the intersection design (Johansson, 2008).

In this study, four trajectory datasets were generated by the social force model under four scenarios. Each scenario is differentiated by the design of the intersection to examine evacuation performance by the proposed trajectory data-mining scheme (Figure 101). Scenario 1 is the base scenario, where pedestrians evacuate from North, West, and South corridors to the East exit on a simple four-way intersection. In Scenario 2, North-East and South-East corners of a corridor are smoothed by rounding right angle corners (Figure 101, top-right), which encourage pedestrians to make smoother turns. Scenario 3 and 4 have the same

street designs as Scenario 1 and 2 respectively except that three bollards are installed on the East corridor as obstacles (Figure 101, bottom images). These obstacles can be seen in the real-world; for example, at entrances to street festivals and outside shopping malls, and holiday promenades to separate pedestrians and vehicles. These obstacles, which limit available space for pedestrians to walk or avoid each other, potentially generate congestion and evacuation bottlenecks.

The spatial extent of the model was set to 800 in width and 700 in height in the simulation unit length, and one unit length corresponds to 1/30 meters (area width=26.7m, area height=23.3m, corridor width & height=5.0m). A pedestrian is represented as a circle with radius equals 10 (0.33m). Pedestrian's desired velocity v_i^0 is approximately Gaussian distributed with a mean value of 1.3 m/s, which represents pedestrian walks in normal situations (Helbing, Buzna, Johansson, & Werner, 2005), and a standard deviation of 0.1 m/s. To determine waypoints for pedestrians, waypoint zones (size: width=10, height=5) were manually introduced and each of pedestrian randomly picks one waypoint in the zone. For pedestrians in the North and South corridors, the x-coordinate of the final destination is the East boundary of the simulation area and y-coordinate is determined by adding a random perturbation value from the y-coordinate of the waypoint. The destination point for pedestrians evacuating from the West corridor is set to the East boundary for x-coordinate and the center of the corridor for y-coordinate. To seed simulation runs, 40 pedestrians were randomly distributed in three starting zones (Total pedestrians = 120).

For each pedestrian, three-dimensional points (x, y, t) were sampled every 1 second (every 100 frames) to create trajectory data. Figure 102 shows trajectories from four simulation scenarios. Two identifiable differences of two-dimensional trajectories among four scenarios are; 1) smoother curves for pedestrians from North and South in Scenario 2 and 4 caused by the effect of rounded corners, and 2) concentrated paths for pedestrians moving from the West due to the effect of avoiding obstacles. However, it is difficult to see clear differences in terms of crowd behavior and process because temporal information is hidden in these two-dimensional trajectory images.

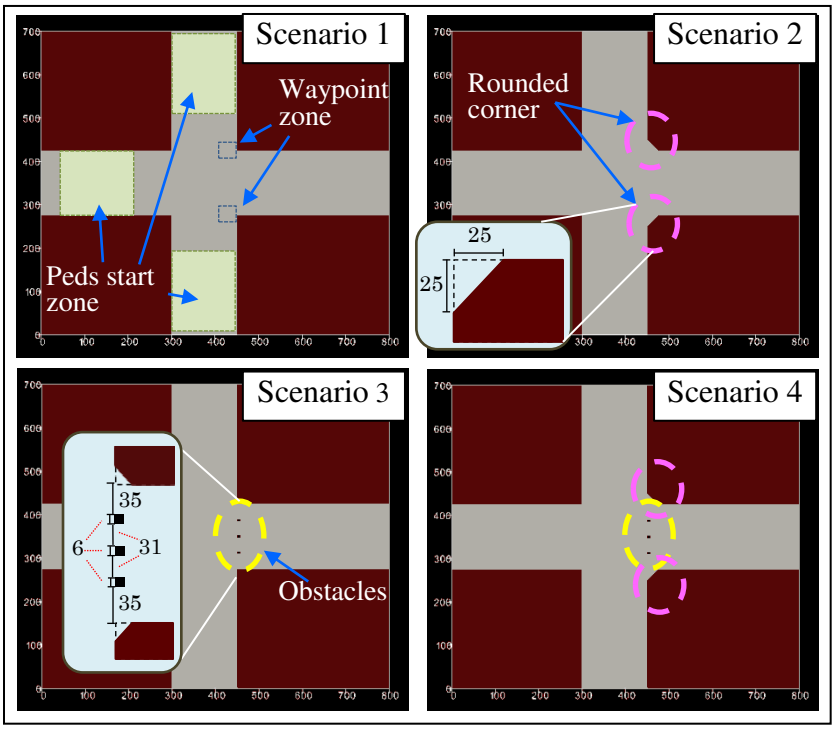

Figure 101. Street designs for four simulation scenarios (numbers represented are in simulation unit length).

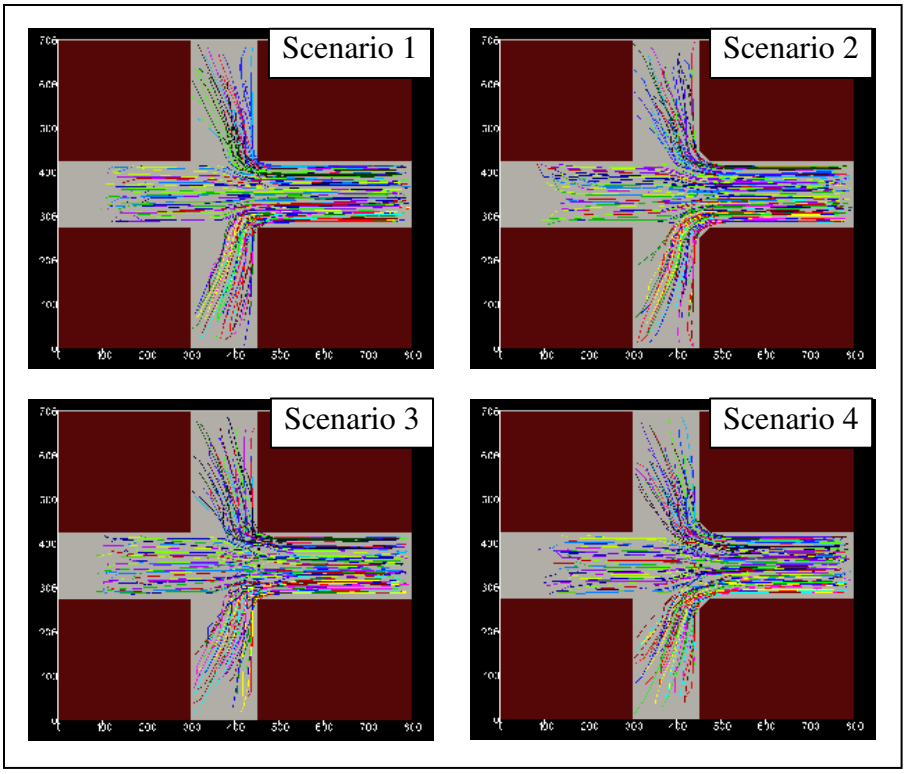

Figure 102. Trajectories of four simulation scenarios.

6.4.2 Evaluating Simulation Scenarios

6.4.2.1 Descriptive statistics of trajectories

Table 24 shows descriptive statistics of motion descriptors of trajectory dataset in four scenarios, whereas Table 25 presents statistics of velocity and acceleration of segments. These statistics summarized pedestrian evacuation behaviors and evacuation efficiency by looking at average values of total egress time, path length, velocity, path sinuosity (straightness index and mean fractal dimension), and directional distribution. Global evacuation efficiency can be measured by the total egress time (max egress time in Table 24), while other descriptors describe relative efficiency. A shorter average path length generally indicates efficient

evacuation; however, in some cases such as multiple flows at a four-way intersection, small detours make a path length longer but decrease the frequency of necessary deceleration, stopping, and avoidance maneuvers so that crowd dynamics become more efficient on average (i.e., roundabout traffic) (Helbing, Molnár, Farkas, & Bolay, 2001). Velocity and acceleration describe general properties of movement relative to a fixed point or to a prior speed. These properties can differentiate motion behaviors; for example, velocity can explain evacuation behaviors like running, walking, and stopping due to bottlenecks, while change in acceleration can describe phase shifts in evacuation behaviors relative to speed. The measurement of sinuosity describes tortuosity, a property of a movement path being tortuous or crooked. Whereas the straightness index looks at sinuosity of a movement path at a global scale, the fractal dimension metric can examine relative sinuosity at different spatial scales. In the case of an emergency evacuation, the movement of a crowd usually is straightforward when pedestrians know the exits and the egress routes; therefore, in such a case, sinuosity tends to indicate paths being straight as compared to, for example, wandering behavior. Circular dispersion describes directional variability in turning angle distribution along a path of a pedestrian. Similar to the argument in sinuosity measurements, circular dispersion tends to be close to 0 (indicating less directional variability because of directed movement behavior) under an emergency evacuation.

As expected, the results show that the mean egress time decreased by rounding rectangle corners at the intersection to encourage evacuees to make smoother turns (Scenario 2 and 4), and increased by inserting obstacles (Scenario

3 and 4) that limited available space for evacuees to avoid other evacuees and obstacles and created congestion. In terms of the total egress time (max egress time), the most efficient intersection design is Scenario2 (19.97 sec) and the worst is Scenario3 (21.93sec). The total egress time is tied in Scenario 1 and 4 (90.96sec) indicating that the positive effect of rounded corners and the negative effect of inserting obstacles are equivalent.

Other descriptors show correspondence to evacuation time. Mean path length, mean circular dispersion, mean fractal dimension, and average acceleration of segments decreased in Scenario 2 and 4, because of shorter travel distance and more directed movement so that evacuation efficiency increased. Increases in average velocity of trajectories and segments, and straightness index describe higher velocity movement and straighter path so that evacuation efficiency also increased in Scenario 2 and 4.

Descriptive statistics can summarize evacuation behaviors in different scenarios, which may be specifically useful for decision makers; however, detailed spatio-temporal information regarding pedestrian behaviors may be hidden under global statistics. Understanding detailed information such as the causes and effects of street designs on crowd behaviors is critical to better design facilities and manage evacuation.

Figure 103 illustrates the visualization of space-time trajectories. To emphasize the temporal effect of crowd behaviors, the value of the time attribute is multiplied by 20. Color and stream tube representation techniques were used to emphasize average velocity variations of segments. Blue and thick tubes denote

low velocity, while red and thin tubes represent high velocity. The STP maps allow us to identify spatio-temporal patterns of crowd evacuation. These clearly highlight spatio-temporal bottlenecks, particularly in Scenario 3 because of the effect of obstacles that limited pedestrians' available space to walk and encouraged congestion. In Scenario 4, on the other hand, as descriptive statistics suggested, the positive effect of rounding corners and the negative effect of obstacles cancel each other out, and evacuation bottlenecks are reduced. Finding this kind of effect is important for evacuation management in the real-world; nevertheless, these representations only show the surface of multiple STPs and much of the movement behaviors are hidden due to occlusion effects created by multiple paths.

Table 24. Descriptive statistics of trajectory's motion descriptors.

	Scenario	Mean	SD	Min	Max
Egress Time (sec)	*1*	13.91	2.82	7.96	20.96
	2	13.51	2.79	7.97	19.97
	3	14.73	3.09	8.93	21.93
	4	13.70	2.98	7.97	20.97
Path Length (unit lengths)	*1*	572.29	67.61	398.81	683.89
	2	565.94	74.57	393.70	699.24
	3	573.55	68.61	428.09	707.14
	4	555.83	69.10	408.17	697.43
Average Velocity (unit lengths / sec)	*1*	41.97	4.66	29.14	52.39
	2	42.65	4.37	30.98	51.94
	3	39.81	4.68	29.11	48.50
	4	41.52	4.99	30.66	51.21
Straight Length (unit lengths)	*1*	506.48	98.33	352.55	680.70
	2	505.80	102.29	345.10	691.74
	3	511.16	99.29	347.81	691.87
	4	502.61	95.41	344.42	687.04
Straightness Index	*1*	0.8795	0.0885	0.7286	0.9996
	2	0.8880	0.0847	0.7386	0.9996
	3	0.8854	0.0841	0.7295	0.9993
	4	0.8985	0.0747	0.7622	0.9989
Circular Dispersion	*1*	0.1285	0.0916	0.0004	0.2772
	2	0.1174	0.0870	0.0004	0.2708
	3	0.1293	0.0883	0.0008	0.3125
	4	0.1096	0.0769	0.0012	0.2531
Fractal Dimension	*1*	1.0149	0.0120	1.0002	1.0649
	2	1.0136	0.0117	1.0001	1.0608
	3	1.0126	0.0091	1.0002	1.0475
	4	1.0118	0.0100	1.0003	1.0535

Table 25. Descriptive statistics of segment's motion descriptors.

	Scenario	*n*	Mean	SD	Min	Max
Average Velocity (unit lengths / sec)	*1*	*1674*	41.10	8.96	3.54	63.89
	2	*1625*	41.87	7.99	7.08	64.38
	3	*1776*	38.84	11.34	2.02	66.53
	4	*1648*	40.55	9.54	2.28	63.97
Average Acceleration (unit velocity / sec)	*1*	*1674*	1.20	7.36	-20.74	37.27
	2	*1625*	0.78	6.13	-19.56	39.43
	3	*1776*	1.76	10.24	-39.65	42.22
	4	*1648*	0.79	7.78	-39.42	47.33

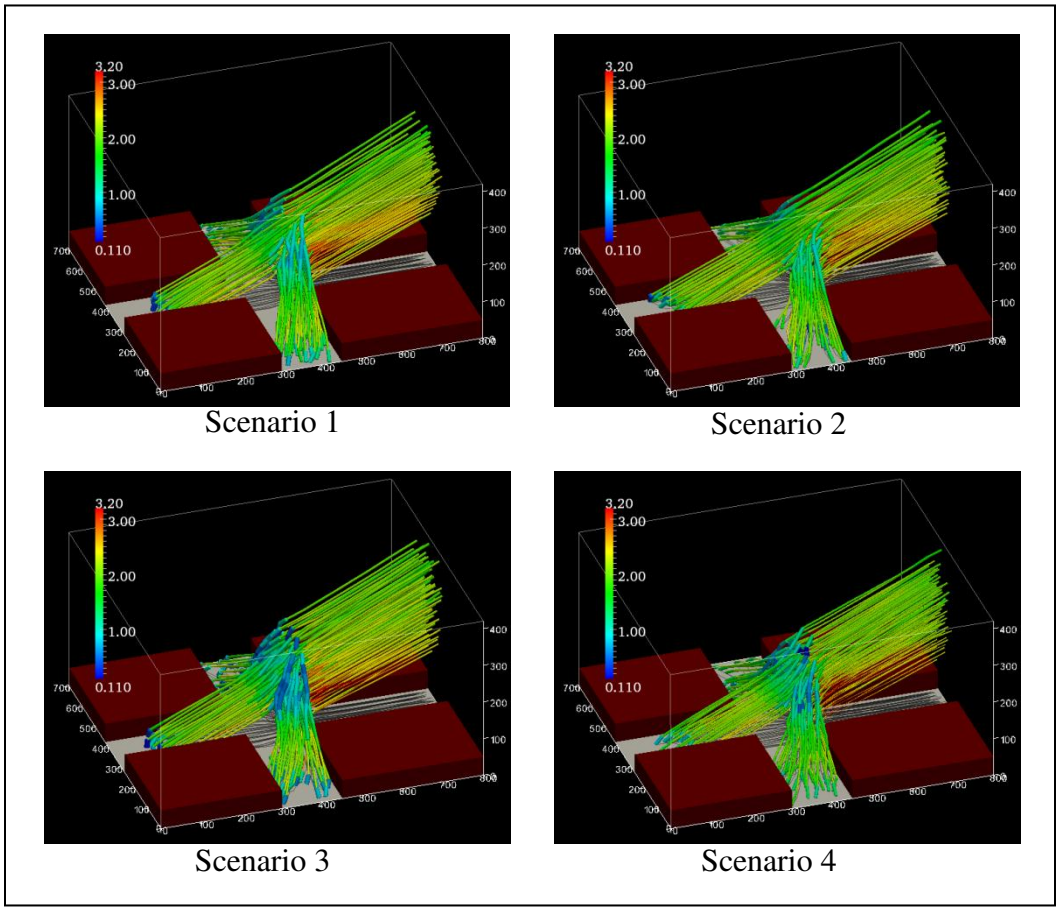

Figure 103. STPs colored by average velocity (unit length/unit time) of segment (Pedestrian evacuation). Clusters of blue color paths describe evacuation bottlenecks. Installing obstacles in Scenario 3 increased bottlenecks around the

intersection corners, while the rounding corner effect in Scenario 4 reduced bottlenecks.

6.4.2.2 Trajectory data-mining: clustering and visualization

Although descriptive statistics of motion descriptors for trajectories can capture important characteristics of crowd dynamics, these consider only basic properties about movement individually from a macroscopic view point. In order to extract complex spatio-temporal patterns of crowd dynamics, trajectory data in each scenario was further investigated using the trajectory data-mining framework.

First of all, trajectory data in each scenario was merged into one trajectory dataset (Figure 105: Top). Using this dataset, the Distance-Threshold approach was used to partition each trajectory because it can differentiate behaviors (between move and stay) along the trajectory. In addition, movements of mobile objects in many situations involve stopping/staying behaviors when people change their behavior. In crowd dynamics, for example, a pedestrian decelerates and ultimately stops to make a sharp turn at an intersection, to avoid collisions with other pedestrians or obstacles, or to wait until traffic jam is cleared up. Identifying these behaviors is important for evacuation management; therefore, the Distance-Threshold is more appropriate as compared to partitioning methodologies purely based on geometrical shapes. The distance threshold value (Th_d) to determine staying behaviors was arbitrarily set to 30 unit length (=1.0m).

For each trajectory partition (sub-trajectory), multi-dimensional vectors of motion descriptors were calculated to characterize the partition trajectory. (This

has a significant advantage because such behaviors are far more complex and cannot be fully described by just a single variable.) Then PC scores of each sub-trajectory for each PC (Eigen value ≥ 1) were calculated, and they were used as a new input for cluster analysis. K-means clustering was run for the input dataset with different k in an arbitrary defined range between 2 and 20. The optimal value of $\hat{k}$ was estimated by applying the gap statistic (see 0). The number of generating reference datasets of a null model, B, was set to 25. As a result, an optimal $\hat{k}$ value 8 was obtained (Figure 104).

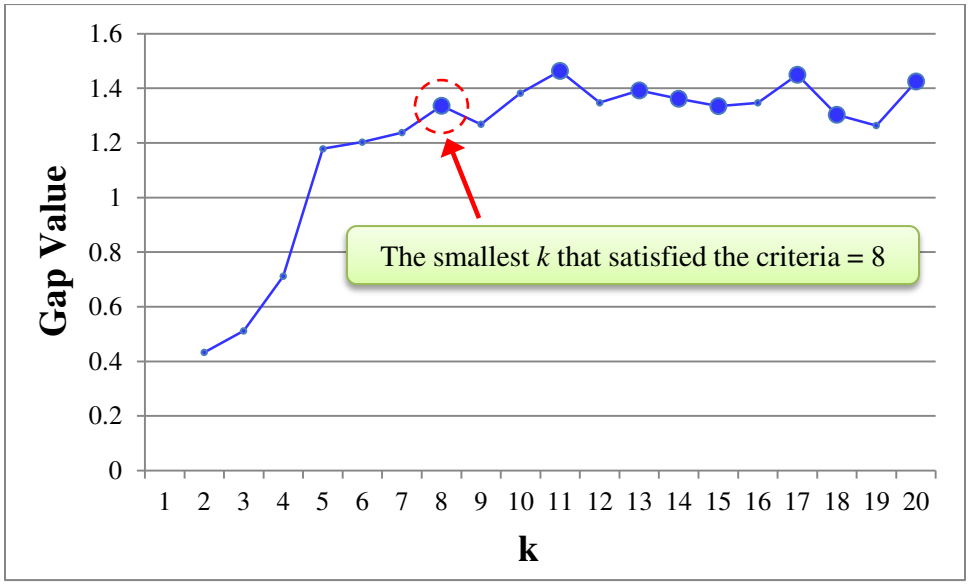

Figure 104. The gap curve for identifying the optimal k value.

The results of trajectory clustering grouped sub-trajectories into 8 behavioral clusters based on multiple motion descriptors (Figure 105, bottom images). The extracted behavioral clusters describe collective movement

behaviors in the simulated evacuation dynamics. In summary, these clusters show smooth and continuous movement (Cluster 3, 4, and 7), clogging (Cluster 6 and 8), slow movement (Cluster 5), and evacuation dynamics fragmented by clogging and slow movement (Cluster 1 and 2). Based on these extracted behavioral clusters, the effects of different designs for intersections on evacuation dynamics are evaluated.

The bottom images of Figure 105 show the result of trajectory partitioning and clustering with k=8. As the result of trajectory partition, Cluster 6 and 8 were classified as *STAY*, describing staying behavior and very slow movement because both distances of each segment in a sub-trajectory and the overall distance of the sub-trajectory were less than the distance threshold value (Th_d). Sub-trajectories of Cluster 5 were initially classified as *STAY*; however, these were re-classified as *Move* because the overall distance of each sub-trajectory exceeded Th_d describing the *SLOW* movement (see section 0). These behaviors of *STAY* and *SLOW* clusters are the indicator of low evacuation efficiency that explains jamming and clogging behaviors creating evacuation bottlenecks. Other clusters (Cluster 1, 2, 3, 4, and 7) were classified as *Move*, and these sub-trajectories can be obtained from partitioning entire paths by sub-trajectories of Cluster 5, 6, and 8.

Figure 106 presents the cluster profile; the vertical axis denotes cluster ID and the horizontal axis shows the average of a normalized value of independent variables within a cluster. This quantitatively describes detailed movement characteristics of clustered sub-trajectories by multiple motion descriptors. Cluster 6 and 8 are identified as a collective staying behavior or very slow

movement described by short travel length and low velocity. The major difference is the directionality of movement. While Cluster 6 represents vertical movement (i.e., movement from North to South or vice-versa), Cluster 8 is horizontal (i.e., movement from West to East). Two behaviors are observed in these clusters. The first is very slow movements observed at the evacuees' starting locations. This explains evacuees' initiation of their body movement. The second behavior is clogging. Cluster 6 describes a clogging behavior, which is created because evacuees from North and South need to make turns so that they decelerate. The deceleration is further propagated back through crowd and created clogging and congestion. This cluster represents a general behavior since it is observed in all scenarios. Cluster 8 also represents clogging behaviors in the middle area of the intersection, and that could be due to obstacles and/or congestion created by Cluster 6. Cluster 8 is particularly observed in Scenarios with obstacles (3 and 4). Cluster 1 and 5 are both approaching to corners from North and South corridors; however, Cluster 5 represents slow movement near the intersection corners describing clogging behavior specifically caused by deceleration for making turns, whereas Cluster 1 represents movements with moderate velocity and longer path length on North and South corridors. Cluster 1 is also a negative indicator for evacuation performance, explained by the fact that its movement is terminated at the intersection due to clogging or slow movement behaviors since trajectories of Cluster 1 are partitioned at the intersection. Cluster 2 and 7 both have straight paths, but Cluster 7 is a long continuous path, while Cluster 2 is fragmented and has a shorter path length. This indicates that Cluster 7 represents smooth

evacuation for evacuees from the West corridor. On the other hand, similar to Cluster 1, evacuees with Cluster 2 went through bottlenecks so that their movement was fragmented by clogging or slow movement behavior. Sub-trajectories of Cluster 3 and 4 both have similar movement characteristics. Both represent evacuees from North and South, whose movements are continuous with high velocity indicating efficient evacuation. The major difference identified is the initial position of evacuees in North and South corridors. In Cluster 3, the initial position of evacuees is closer to the intersection, and thus, their path length is shorter.

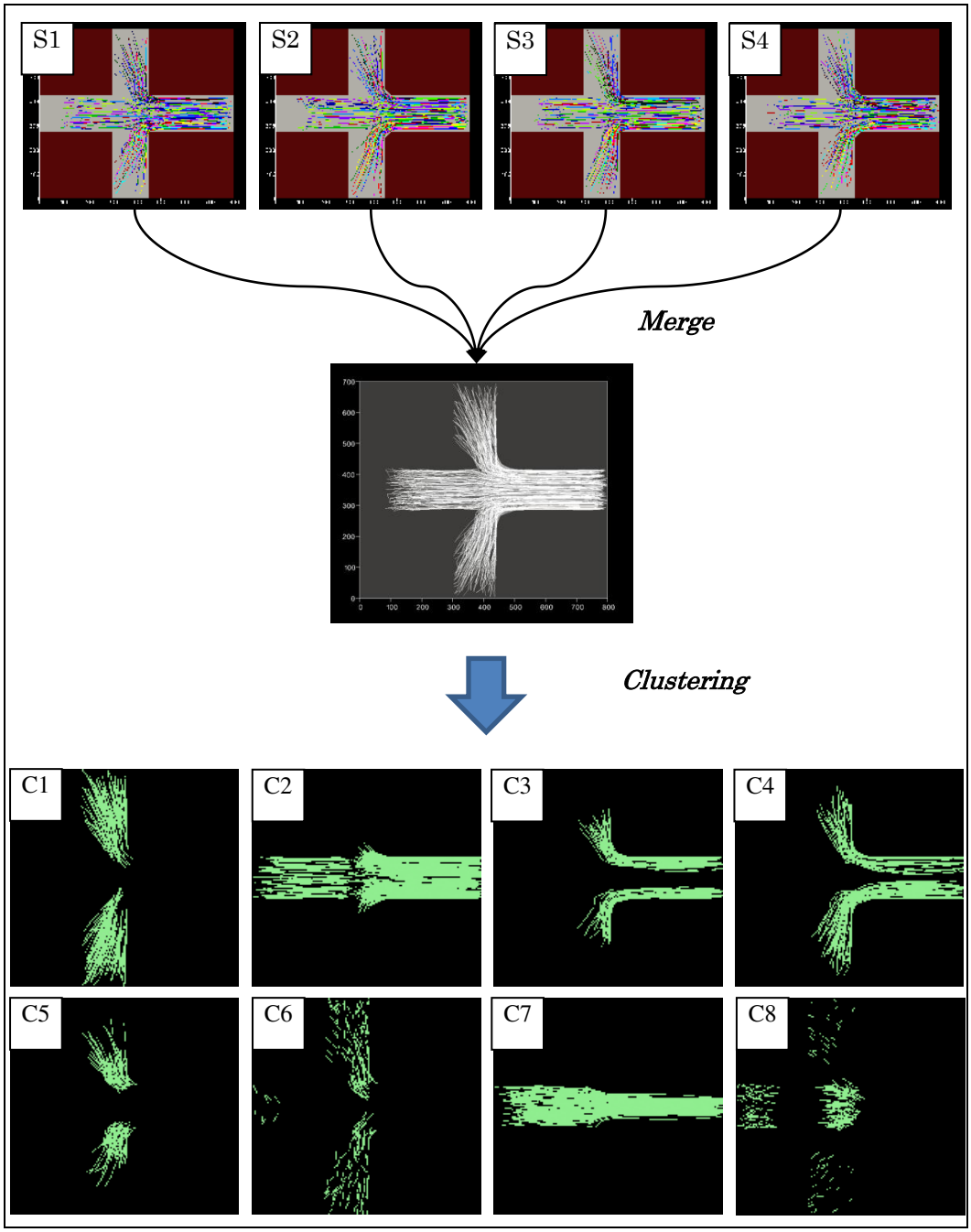

Figure 105. Trajectory clustering framework and result.

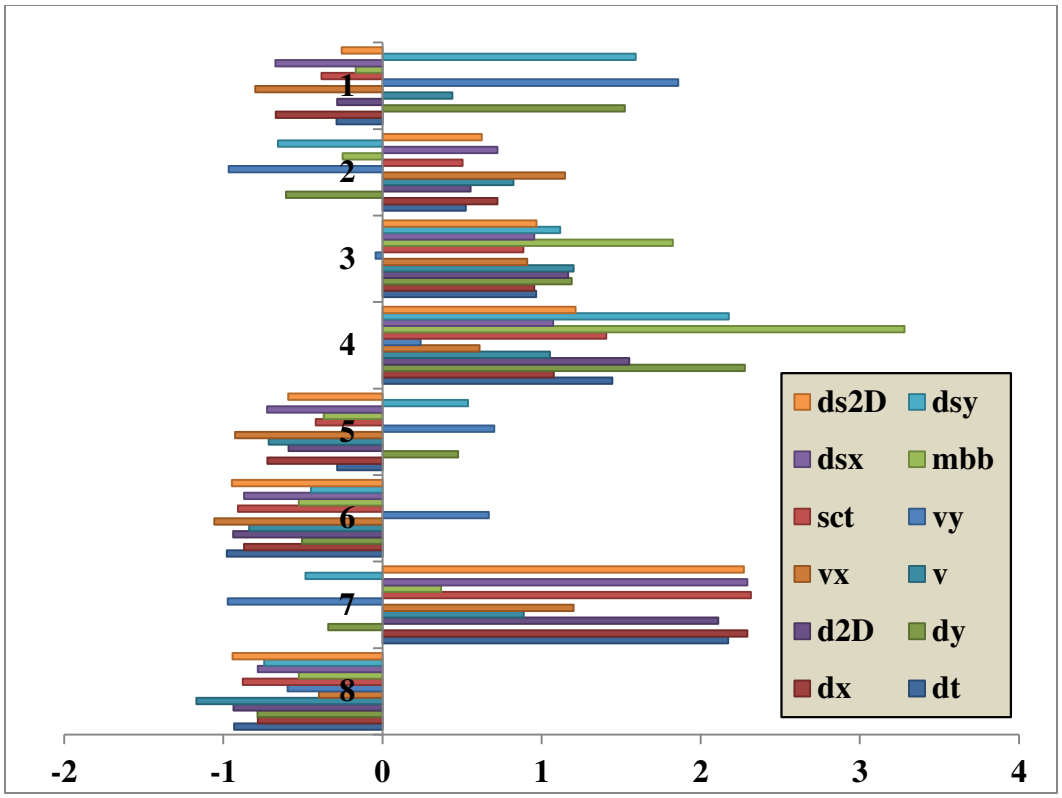

Figure 106. Cluster profiles for pedestrian evacuation simulation.

To investigate extracted behavioral clusters and to compare similarity and dissimilarity in different simulation scenarios, I compared the cluster distribution among four simulation scenarios globally, temporally, and spatio-temporally. The global approach compares the overall movement behaviors based on summarized behavioral cluster distribution, which captures general behavioral differences among scenarios. The temporal approach further investigates and compares the distribution of behavioral cluster through time, which enables us to examine when a particular behavior occurred, if there is any time lag in the occurrence of a particular behavior in different scenarios, and if so why such a lag is observed. Finally, the spatio-temporal approach investigates and compares the distribution

through space and time, which helps in answering questions such as when and where a particular movement behavior is observed in one scenario and not in others, and how and why such different movement behaviors appear.

6.4.2.2.1 Global analysis of behavioral cluster

To identify the global properties of behavioral clusters in four scenarios, the proportions of evacuees' cumulative time within clusters in each scenario are illustrated in Figure 107. This summarizes the overall movement behaviors and allows comparison of behavioral differences among four simulation scenarios. The comparison between Scenario 1 and 2 describes the influence of rounded corners. The significant differences are the decrease in Cluster 2 (-5.85%) and the increase in Cluster 7 (+4.94%) in Scenario 2. This describes the effect of rounding the right angle corners, which caused the number of successful evacuees with smooth and continuous paths to increase (Cluster 7), while the number of unsuccessful evacuees described by fragmented paths decreased (Cluster 2). Another difference identified is the increase in Cluster 3 (+3.60%) in Scenario 2. This suggests that the rounded corners effect increased the number of successful evacuees from North and South, who were encouraged to make smoother turns.

The comparison between Scenario 1 and 3 describes the influence of installing bollards as obstacles. The differences are increases in Cluster 2 (+17.25%), Cluster 5 (+3.26%), and Cluster 8 (+5.63%) and decreases in Cluster 4 (-8.15%) and Cluster 7 (-17.33%) in Scenario 3. This shows that the effect of obstacles increased clogging behavior (Cluster 8), slow movement (Cluster 5),

and fragmented paths (Cluster 2) caused by bottlenecks, decreased successful evacuation dynamics (Cluster 4 and 7), and thus the evacuation efficiency decreased.

The distribution of behavioral cluster in Scenario 4 reasonably explains the mixed effects of rounded corners and obstacles. Comparing the cluster distribution to Scenario 3, Scenario 4 improved evacuation efficiency by reducing clogging and slow movement behaviors (Cluster 5: -1.73%, Cluster 6: -1.49%, Cluster 8: -2.15%) and fragmented paths (Cluster 2: -2.01%), and by increasing successful evacuation dynamics (Cluster 3: +1.78%, Cluster 4: +8.25%).

As compared to the comparison between the base scenario and obstacles scenarios (Scenario 3 and 4), the amount of behavioral difference is small in the comparison between the base scenario and the one with rounded corners (Scenario 2). This indicates that the influence of obstacles on movement behaviors in evacuation dynamics is larger than that of rounded corners.

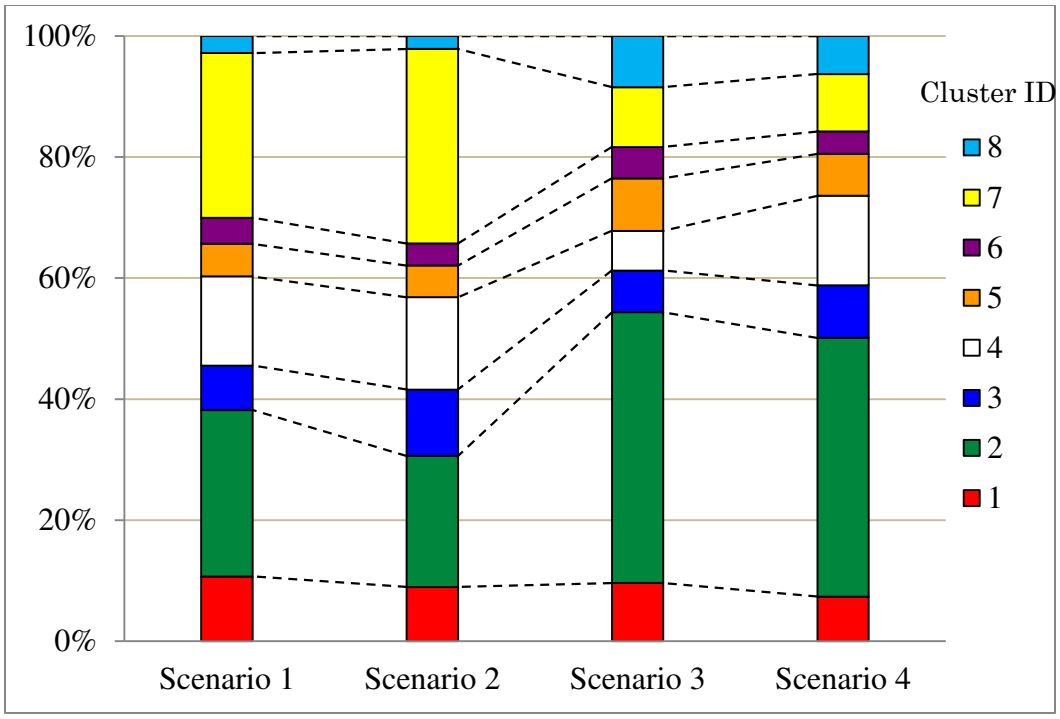

Figure 107. Proportion of clusters in each scenario.

6.4.2.2.2 Temporal analysis

To examine temporal differences in the occurrence of behavioral clusters in different scenarios, three types of figures were created. Focusing on the individual scale, the first figure (Figure 108) shows maps of cluster distribution of individuals through time for each simulation scenario. The vertical axis represents each individual trajectory ID. The IDs equal to 1 to 40 are evacuees from the North corridor, 41 to 80 are those from the West corridor, and 81 to 120 are those from the South corridor respectively. The horizontal axis is the simulation time (unit: second). Each pixel in the images represents a cluster ID at a certain simulation time-step. The major difference identified is the behavioral cluster distribution of evacuees from the West between Scenarios 1 and 2 and Scenarios

3 and 4. In Scenarios 3 and 4, the dynamics of evacuees from the West corridor is fragmented (Cluster 2) by clogging behavior (Cluster 8) in the middle of evacuation. This is the result of installing obstacles at the intersection in those scenarios. Other differences can be found in these maps; however, it is useful to quantitatively distinguish the difference between scenarios. Therefore, from these individual based maps, stacked bar graphs are created to summarize the occurrence of each behavioral cluster through time (Figure 109 to Figure 112). The vertical axis represents the total number of evacuees for each cluster, and the horizontal axis represents time. Whereas Figure 108 describes behavioral cluster distribution of individuals through time, these provide a summary of behavioral cluster dynamics in each scenario. In order to clarify the difference between the base scenario and others, the amount of cluster occurrences at a certain time in one scenario (Scenario 2, 3, or 4) is subtracted by that in the base scenario (Figure 113 to Figure 115). In these graphs, the positive value of a cluster at a time represents that the first scenario has more of that cluster than the second scenario, and the negative value is vice-versa.

Scenario 1 (base) and 2 (rounded corners) (Figure 109 and Figure 110) present very similar dynamics of cluster distribution. This suggests that, despite the installation of rounded corners, the behavioral structures of crowd dynamics between the two scenarios were not changed significantly. This matches the results from global comparison, suggesting that the effect of rounded corners is less influential for evacuation dynamics at each time step than the effect of obstacles. The difference is also captured in Figure 113. In the figure, the initial

behavioral difference at time 1 is identified. This can be explained by the effect of initial positions of evacuees, which are randomly determined within the predefined zones. The random effect varies the initial spatial configuration of evacuees so that their initial behavior also differs. Other differences identified include more occurrences of Cluster 3 (smooth & continuous) and Cluster 7 (smooth & continuous) and less occurrences of and Cluster 2 (fragmented path) in Scenario 2. Cluster 3 represents successful evacuees from North and South corridors because their initial position is closer to the intersection so that they can evacuate smoothly without being involved with congestion. Rounded corners lead to an increase in successful evacuees (Cluster 3) from North and South corridors by encouraging them to make smooth turns at the early stage of evacuation. In addition, the effect in Scenario 2 persisted longer than in Scenario 1 (Scenario 1 = 12 sec vs. Scenario 2 = 11 sec) (Figure 109 and Figure 110). Moreover, there is a time lag of Cluster 5 between two scenarios. Cluster 5 represents clogging behavior due to deceleration for turns. In Scenario 2, clogging behavior occurred earlier than in Scenario 1, which is reasonable because distances to the edge of corners are shortened for evacuees from the North and South. In Scenario 2, this early clogging behavior reduced the same behavior later (Figure 113) when evacuees from the West entered the intersection, which ultimately influenced the increase of successful evacuees from the West (Cluster 7) and decreased of unsuccessful evacuees (Cluster 2).

Between Scenario 1 (base) and 3 (bollards), Figure 109 and Figure 111 show significant differences of behavioral cluster distribution dynamics. In

Scenario 3, the number of successful evacuees from the West (Cluster 7) and North and South (Cluster 4) were decreased, whereas fragmented paths (Cluster 2) in all stages, clogging on North and South corridors at the early stage (Cluster 5), and clogging at the intersection (Cluster 8) in the middle of evacuation were increased. This indicates that the installation of bollards narrowed the space for evacuees from North and South to make their turns so that the clogging behavior emerged at the early stage (Cluster 5), which ultimately reduced the number of successful evacuees (Cluster 4 and 7). In addition, when three flows merged at the intersection, another clogging behavior emerged (Cluster 8) because of the limited flow capacity due to obstacles. This behavior further reduced successful evacuees especially from the West corridor (Cluster 7) and increased fragmented paths (Cluster 2) in the late stage (Figure 114).

As with the global approach, the temporal distribution of behavioral cluster in Scenario 4 described mixed effects of rounded corners and obstacles. Between Scenario 3 and 4, Figure 111 and Figure 112 show similar behavioral distribution dynamics, but the dynamics of Scenario 4 take the effects of rounded corners into account. Besides the initial random effect, the effects include early clogging behavior (Cluster 5) and an increase in successful evacuees (Cluster 3, 4 and 7). Figure 115 compares the temporal behavioral cluster distribution between Scenario 1 and 4, and it shows similar dynamics with Figure 114, but incorporated the effects described above.

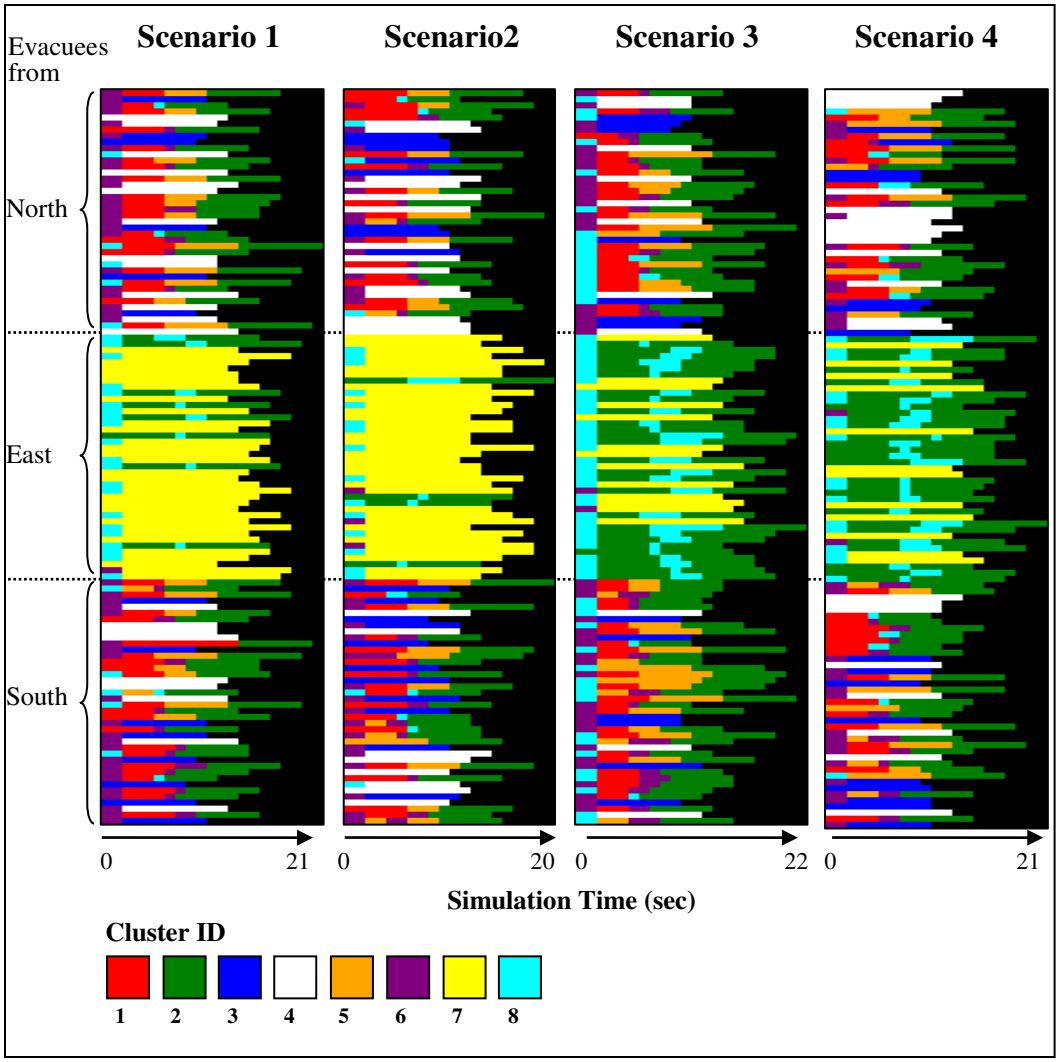

Figure 108. Temporal cluster distribution of evacuees in each scenario.

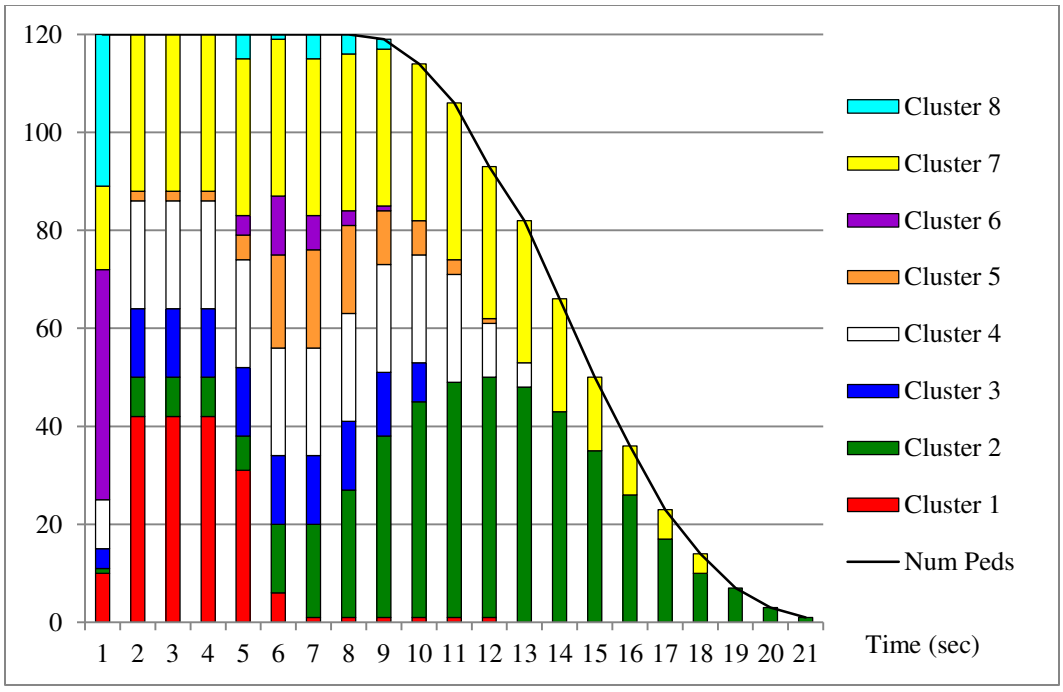

Figure 109. Summarized temporal cluster distribution in Scenario 1.

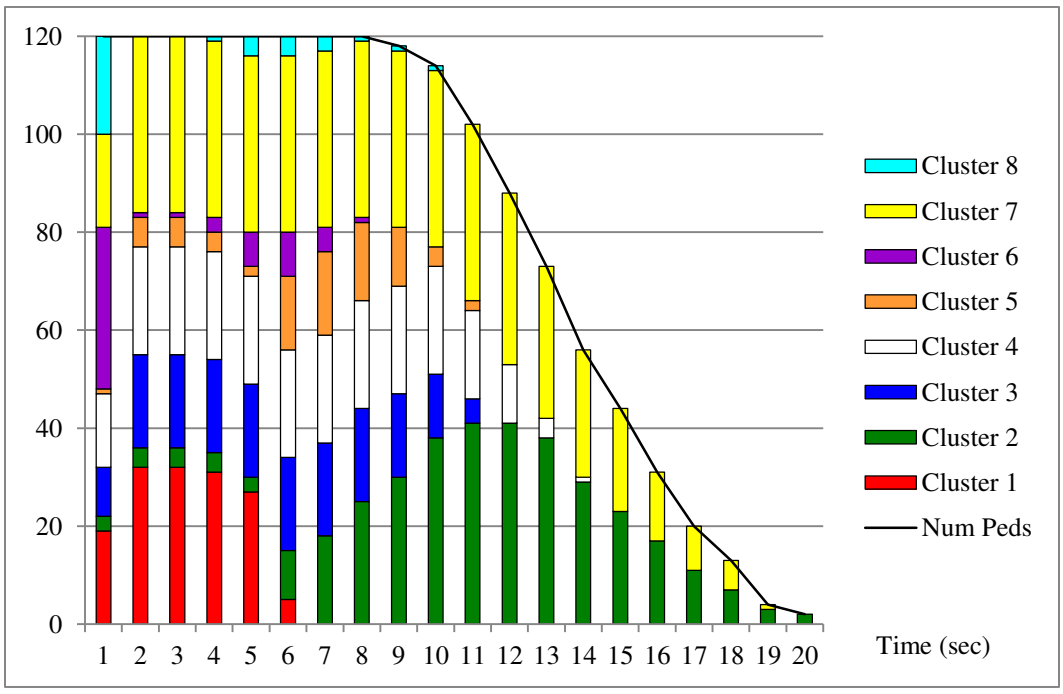

Figure 110. Summarized temporal cluster distribution in Scenario 2.

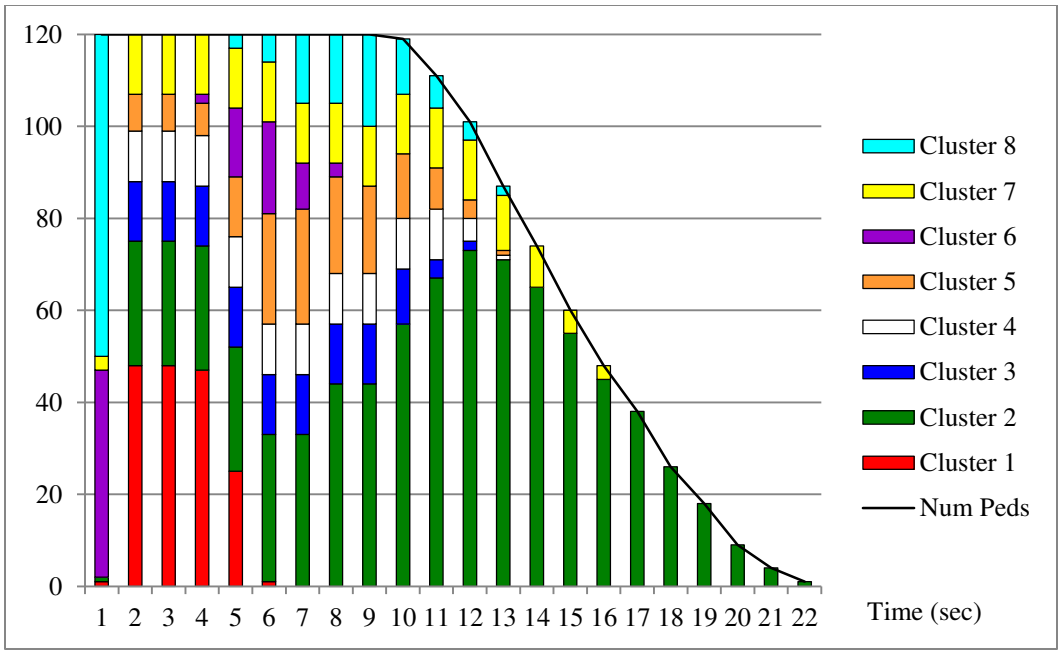

Figure 111. Summarized temporal cluster distribution in Scenario 3.

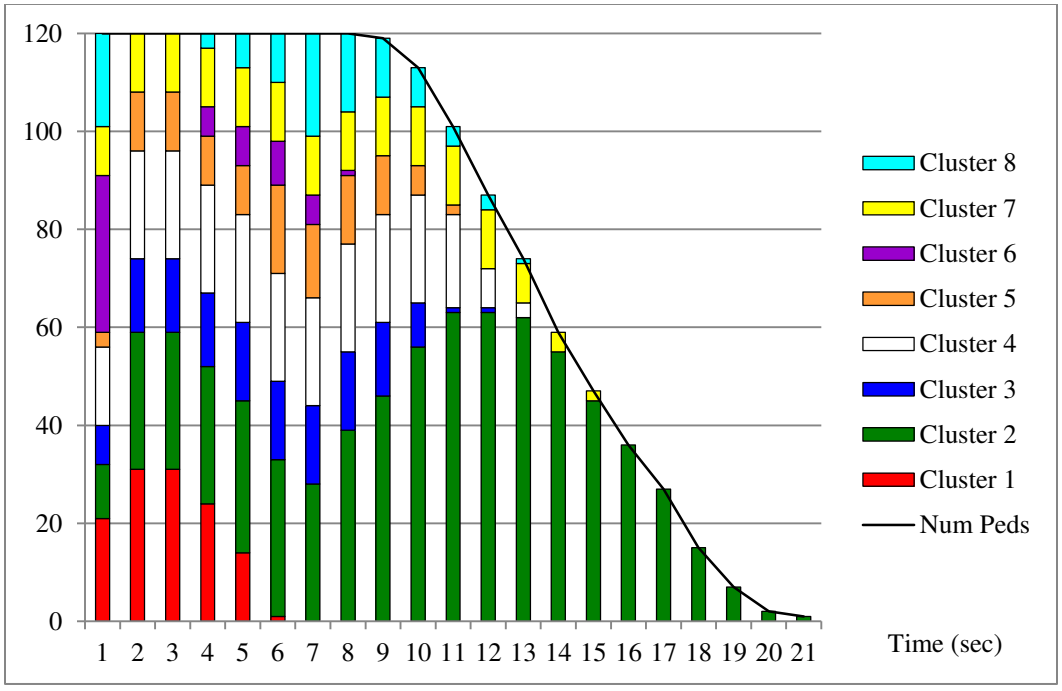

Figure 112. Summarized temporal cluster distribution in Scenario 4.

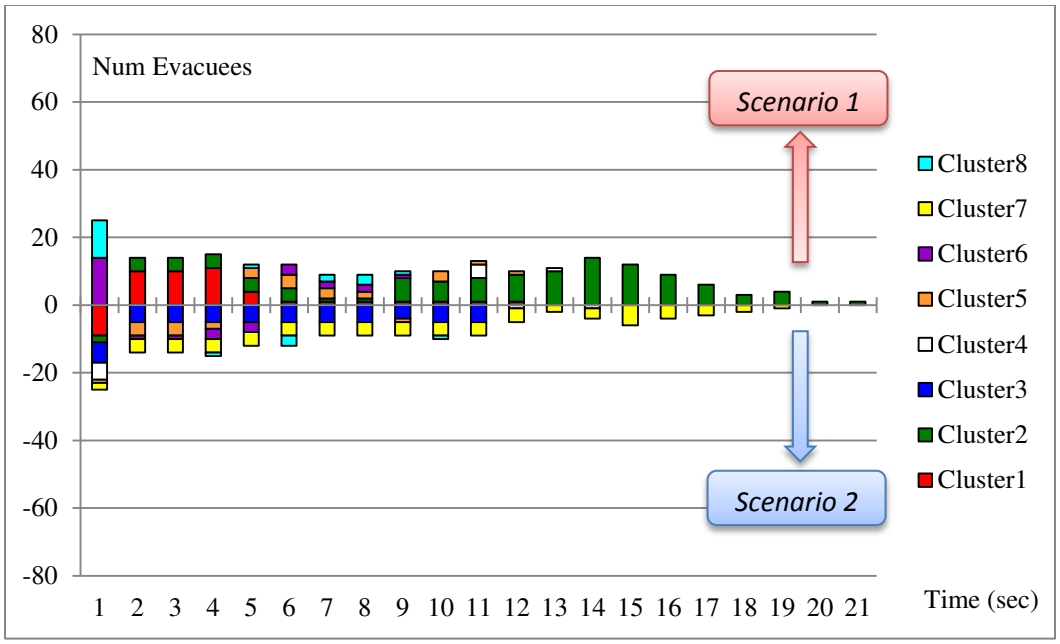

Figure 113. Comparison of summarized temporal cluster distribution between Scenario 1 and 2.

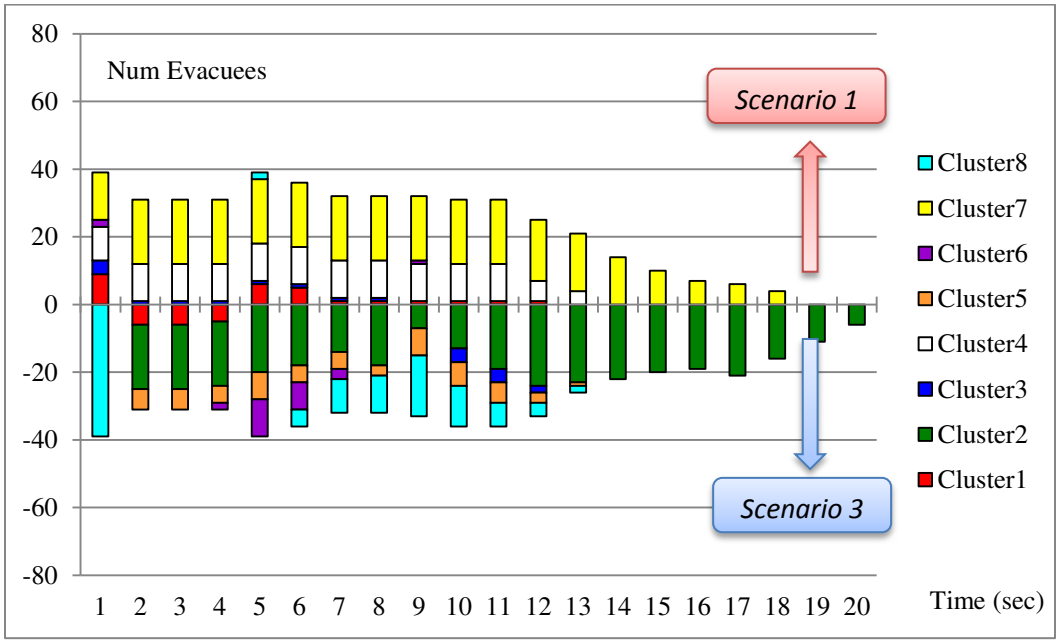

Figure 114. Comparison of summarized temporal cluster distribution between Scenario 1 and 3.

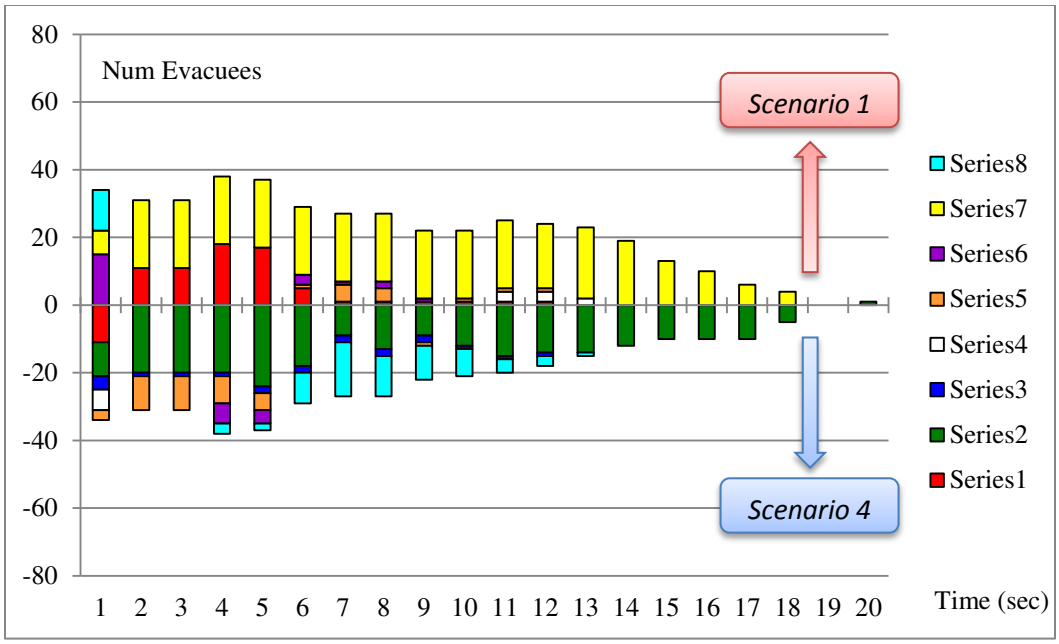

Figure 115. Comparison of summarized temporal cluster distribution between Scenario 1 and 4.

6.4.2.2.3 Spatio-temporal analysis

To explore and compare the spatio-temporal distribution of behavioral clusters, three techniques of time geography were used; STPs, STKDE, and STKDE with map algebra. Figure 116 to Figure 123 show STPs for 8 behavioral clusters in each scenario, which provide detailed spatio-temporal movement behaviors and confirm previously described behaviors by global and temporal approaches discussed in this section. Figure 116 shows the STPs of Cluster 1 describing the movement behavior of evacuees from North and South approaching intersection corners with moderate velocity. This behavior is terminated when evacuees reach at the corners of the intersection due to *STAY* or *SLOW* behaviors, such as deceleration to make turns and clogging by overcrowd. Also, this behavior is

observed in all scenarios, which indicates a general pattern for crowd dynamics regardless of different designs of intersection.

Figure 117 shows the STPs of Cluster 2, which have straight paths but are fragmented. Because of the fragmented paths, evacuees with Cluster 2 were involved in some congestion. Obviously, the existence of obstacles prevented West evacuees from the smooth and continuous evacuation found in Scenario 3 and 4. Interestingly, East evacuees, whose initial positions were nearby the walls (i.e., on the outer side of the corridor), only show this behavior (Cluster 2). This suggests that when three flows merged at the intersection, evacuees have a higher risk to be involved with congestion because of inflow from North or South corridors as well as the high density in the middle of intersection. By introducing round corners, the number of evacuees at risk was decreased because the pressure of inflow from North and South was reduced.

Figure 118 and Figure 119 show the STPs of Cluster 3 and 4 respectively. Both represent successful evacuations from North and South with continuous trajectories. The major difference identified is the initial position of evacuees in North and South corridors. In Cluster 3, the initial position of evacuees is closer to the intersection, while that in Cluster 4 is on the middle of corridors. This difference results in the relatively slow velocity of evacuees in Cluster 4 when they were on North or South corridor. The slow velocity can be explained by the initiation of body movement as well as the feedback effect of deceleration, in which the effect of an evacuee's slowing-down at corners in order to make a turn

is propagated to the crowd behind. Ultimately, this effect created congestion around the intersection corners.

Figure 120, Figure 121, and Figure 123 show the STPs of Cluster 5, 6, and 8 respectively. These three clusters represent clogging behavior. Besides the initial slow movement, Cluster 6 describes a clogging behavior created by evacuees from North and South who were affected by the feedback effect of deceleration (from Cluster 3 and 4) and thus involved in congestion. Cluster 5 also describes a clogging behavior created by the effect of three flows merged together in addition to the feedback effect of deceleration. The two effects created a long clogging behavior. Besides the initial slow movement, Cluster 8 represents a clogging behavior at the middle of the intersection, which is particularly observed in scenarios with bollards (Scenario 3 and 4). This reasonably explains the behavior caused by the obstacles. In addition, sub-trajectories in Cluster 8 describe a zipper pattern during evacuees entrance of the bottleneck created by bollards. This pattern occurs when pedestrians alternatively enter the bottleneck, and the behavior further produces a zipper effect, which is a self-organizing phenomena leading to an optimization of the available space and velocity inside the bottleneck (Hoogendoorn & Daamen, 2005; Seyfried, et al. 2007).

Cluster 7 is a long continuous path representing smooth evacuation for evacuees from the West corridor without being involved with congestion (Figure 122). Obviously, the successful evacuation from the West corridor increased in Scenario 2 with rounded corners, but reduced in Scenario 3 and 4 with obstacles.

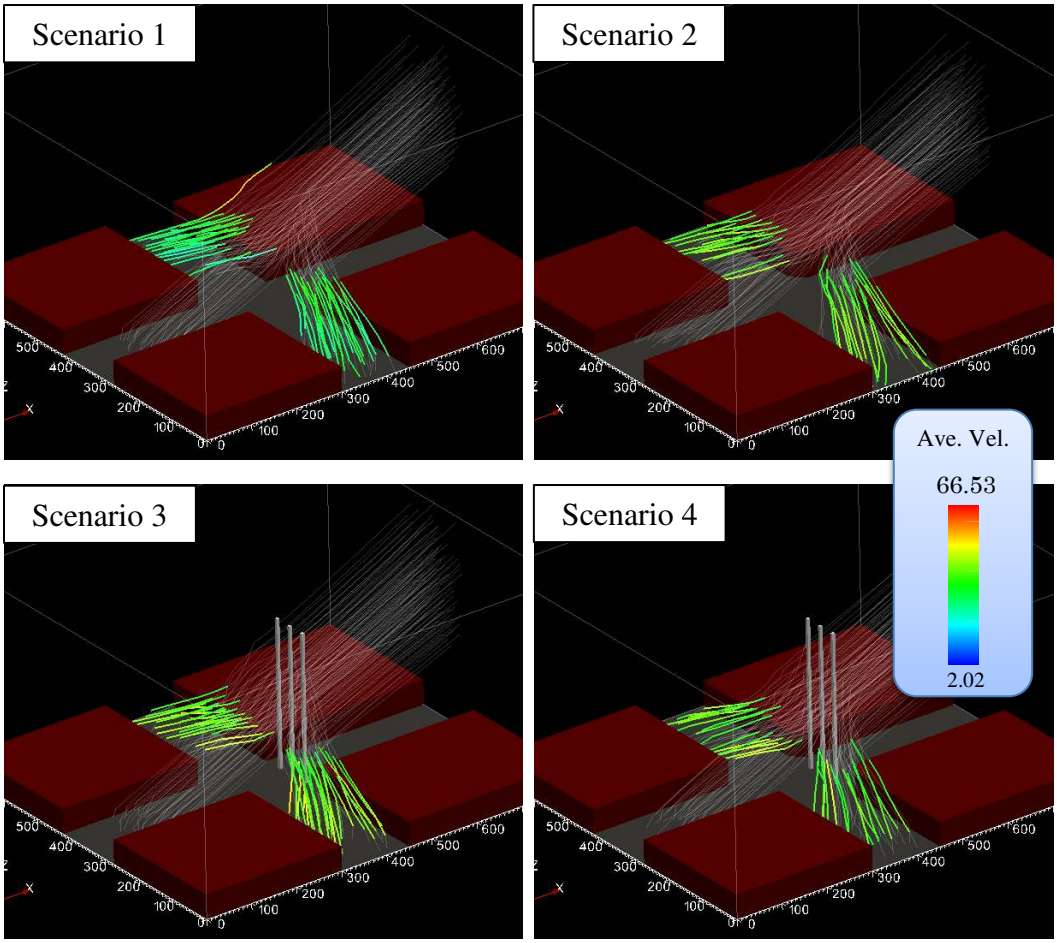

Figure 116. STPs of movements with moderate velocity described by Cluster 1. Color of path represents average segment velocity of a sub-trajectory (unit length / second).

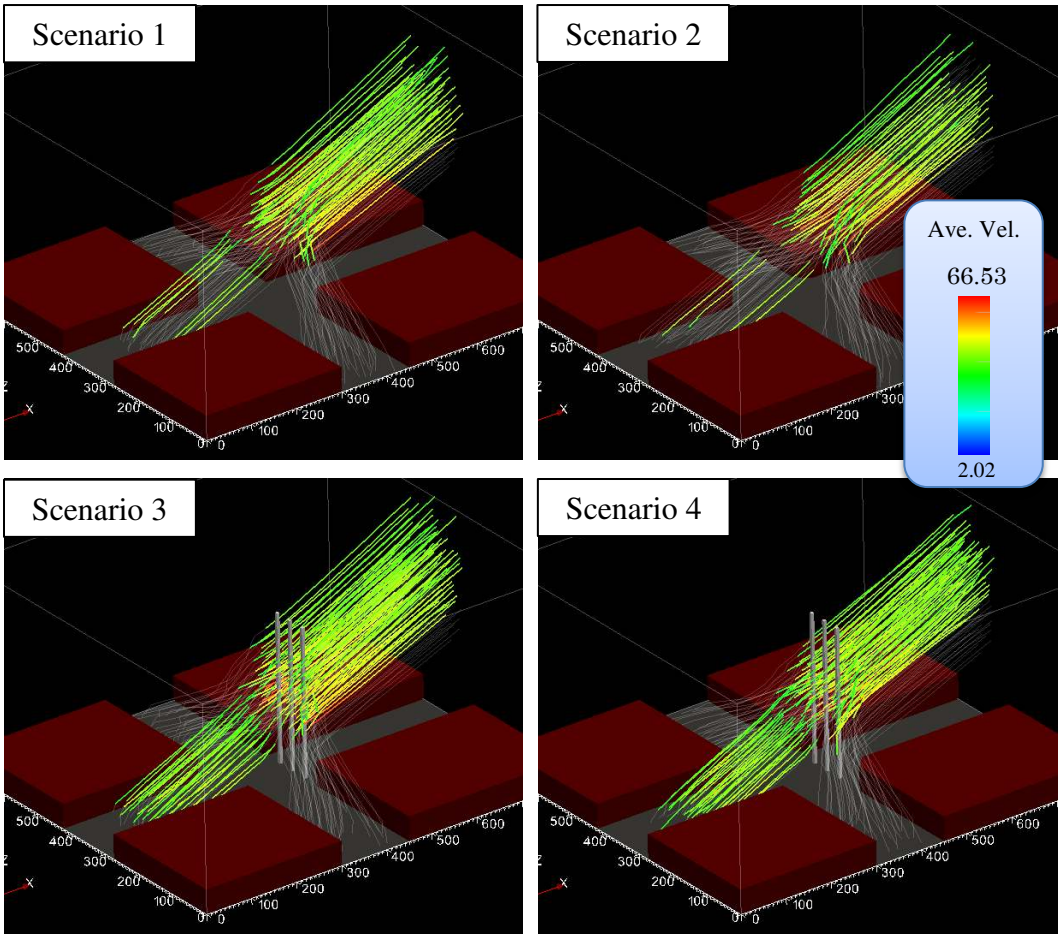

Figure 117. STPs of fragmented paths described by Cluster 2. Color of path represents average segment velocity of a sub-trajectory (unit length / second).

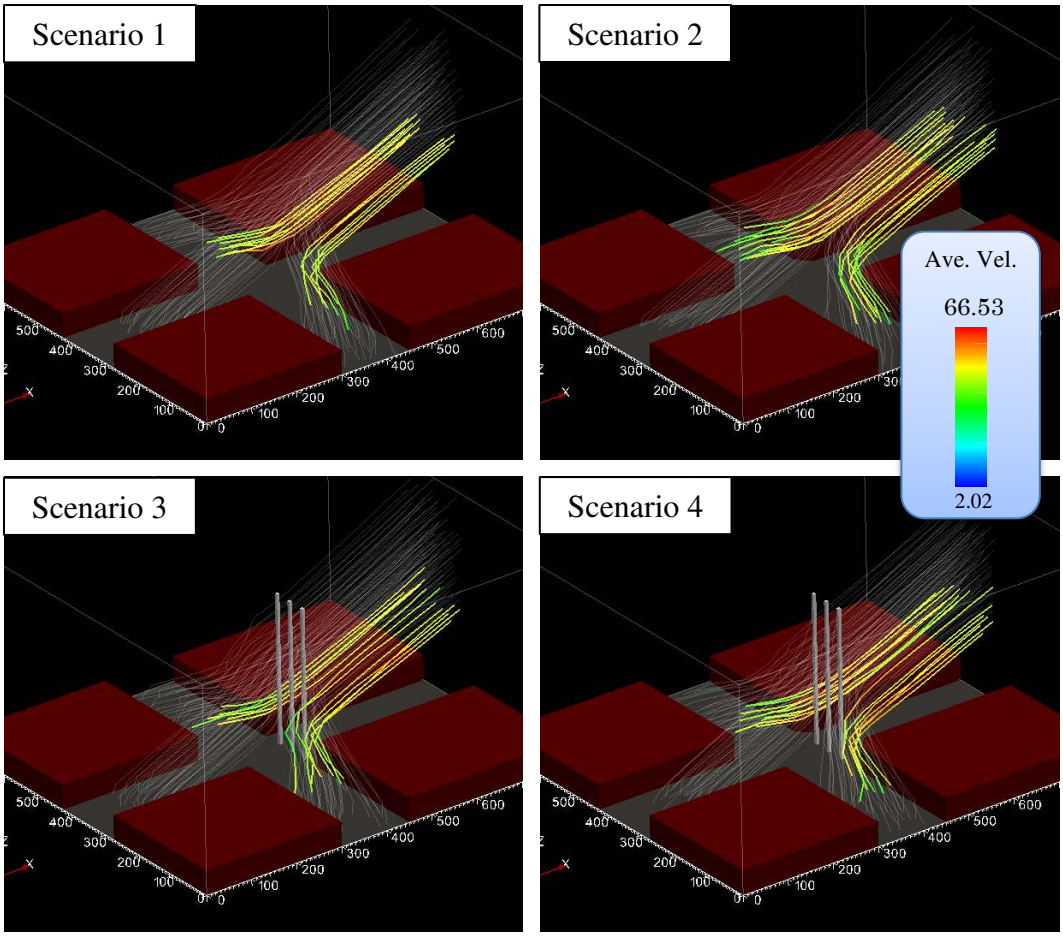

Figure 118. STPs of successful evacuation described by Cluster 3. Color of path represents average segment velocity of a sub-trajectory (unit length / second).

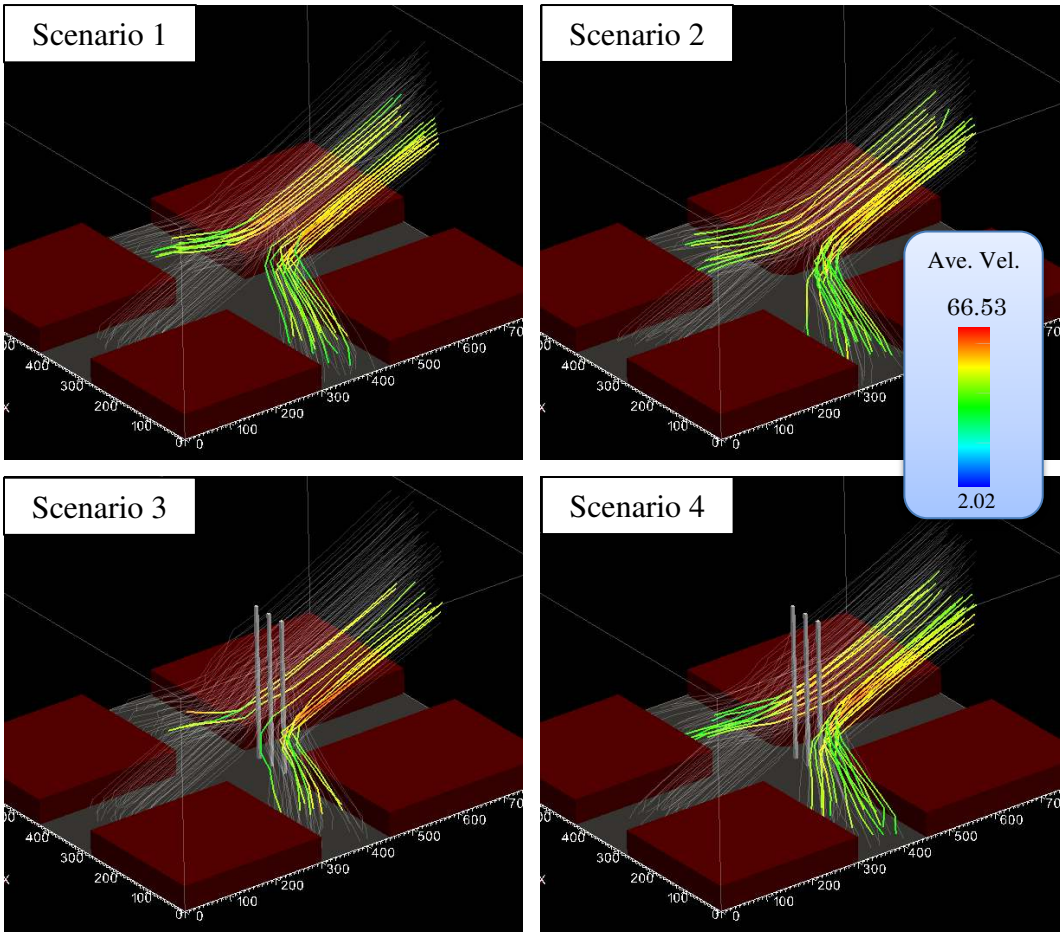

Figure 119. STPs of successful evacuation described by Cluster 4. Color of path represents average segment velocity of a sub-trajectory (unit length / second).

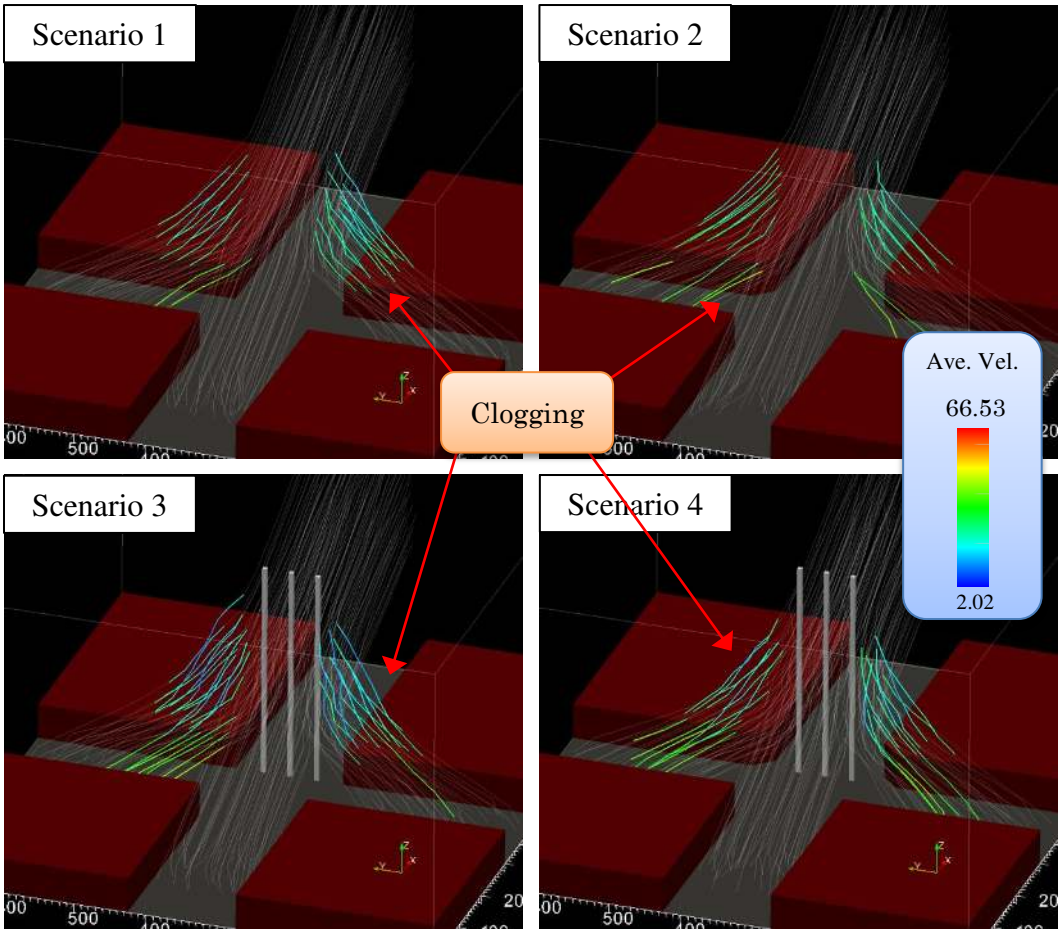

Figure 120. STPs of slow movement described by Cluster 5. Color of path represents average segment velocity of a sub-trajectory (unit length / second).

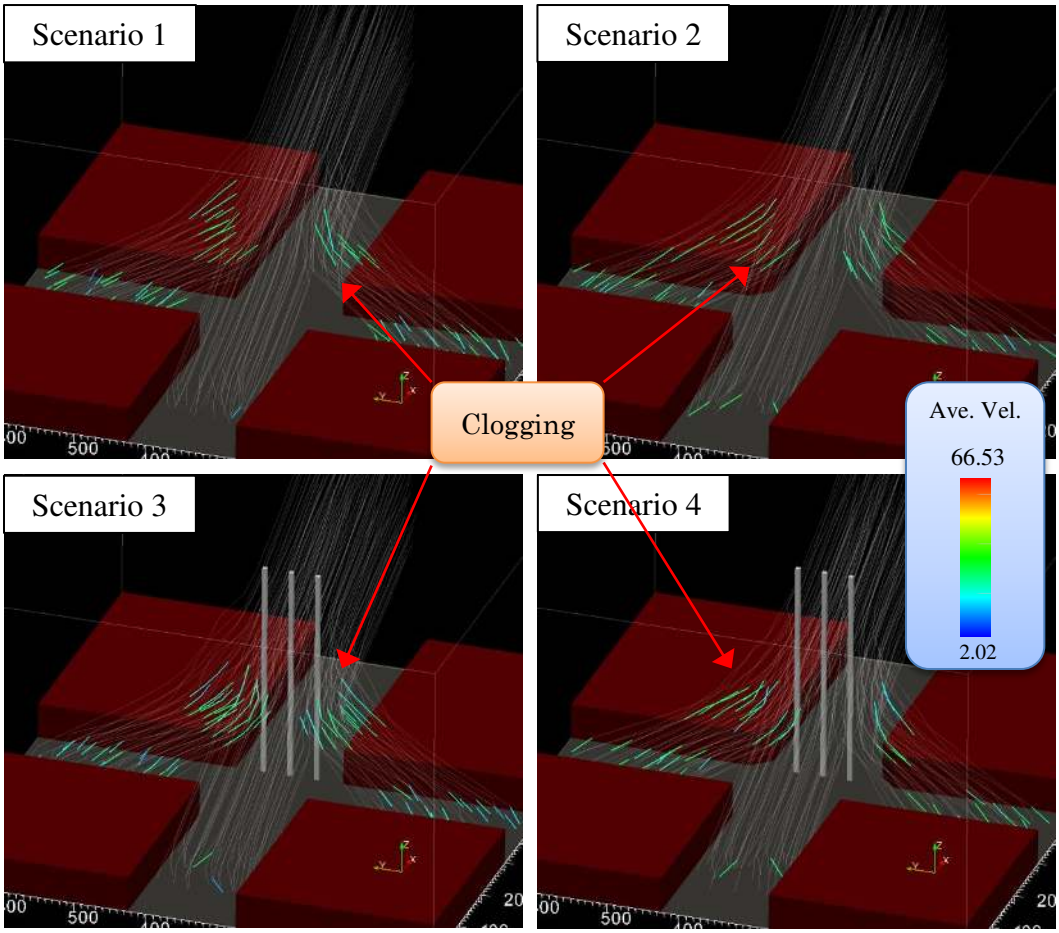

Figure 121. STPs of slow movement described by Cluster 6. Color of path represents average segment velocity of a sub-trajectory (unit length / second).

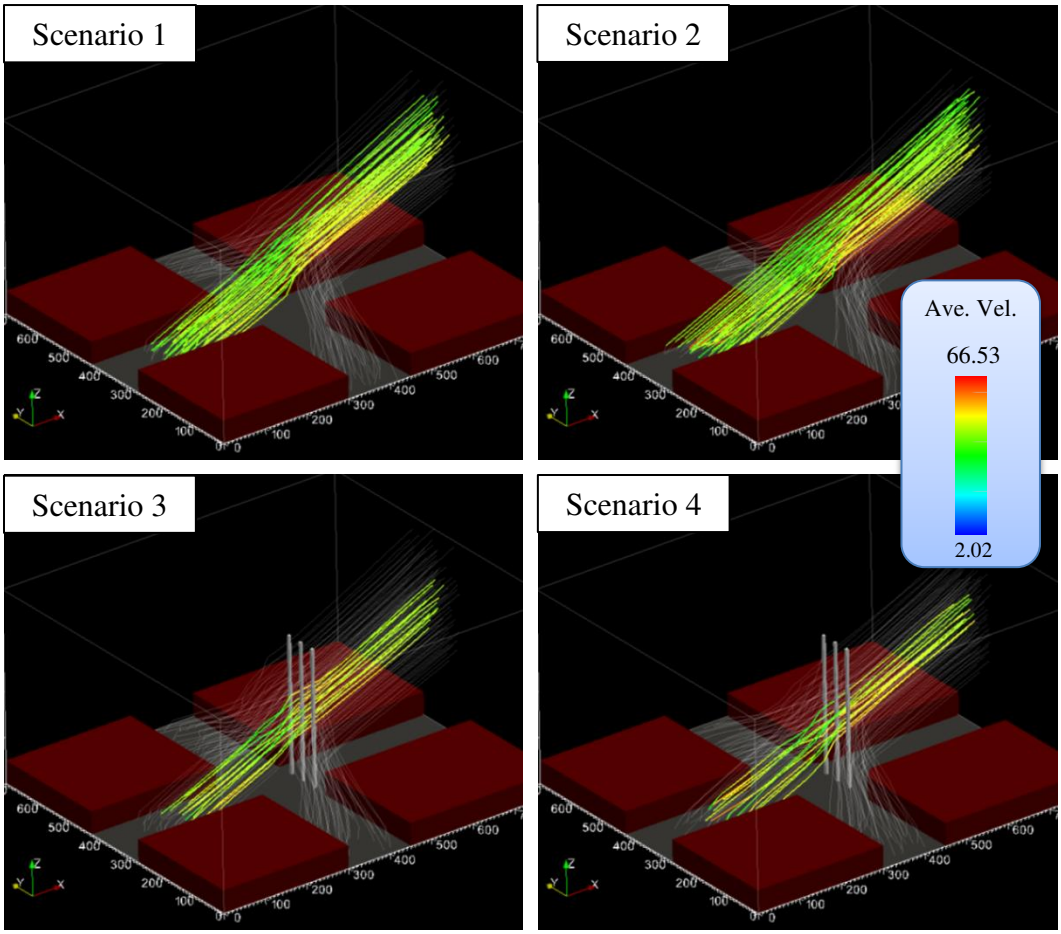

Figure 122. STPs of successful evacuees from the West corridor described by Cluster 7. Color of path represents average segment velocity of a sub-trajectory (unit length / second).

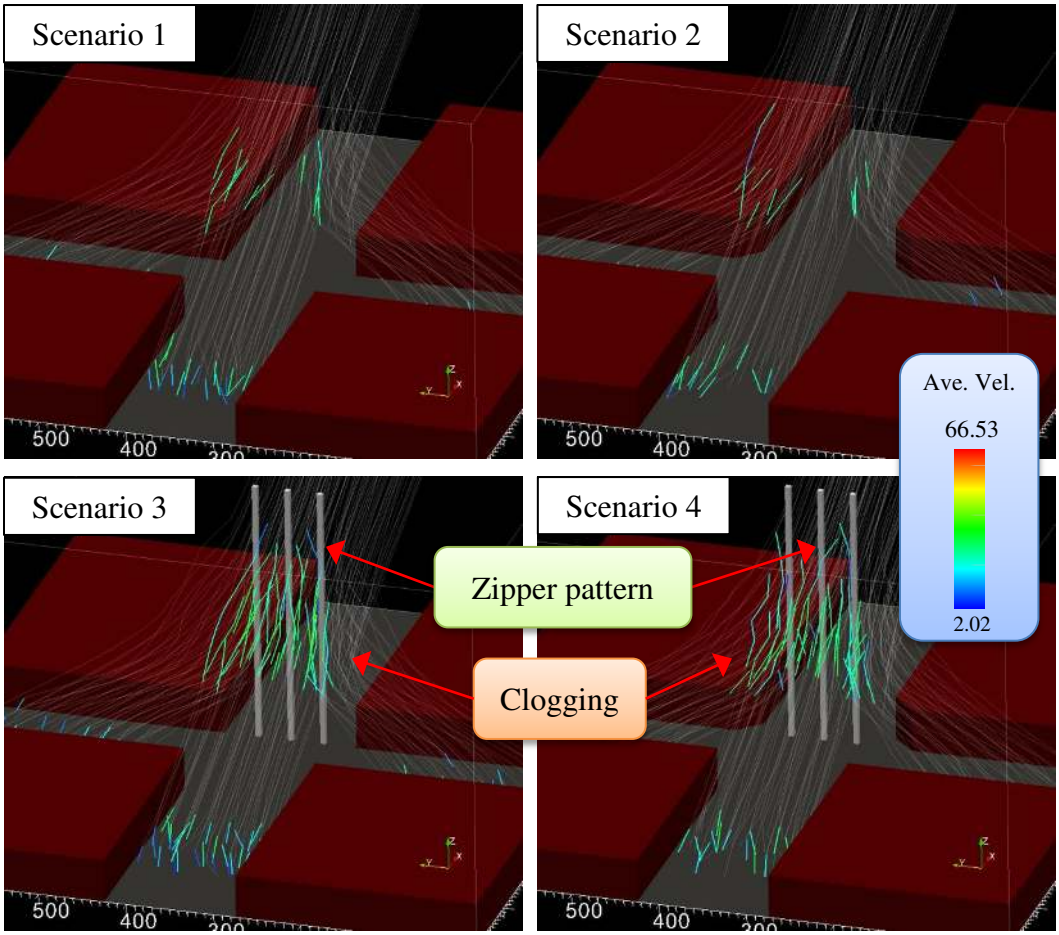

Figure 123. STPs of clogging behavior described by Cluster 8. Color of path represents average segment velocity of a sub-trajectory (unit length / second).

To compare behavioral cluster distribution in space and time between two scenarios, I first applied the STKDE to summarize the distribution pattern from trajectory datasets and then employed 3D Map algebra to calculate and visualize the difference. Figure 124 to Figure 131 illustrate Space-Time line density maps of corresponding cluster ID in each scenario (unit: unit lengths × unit area^{-1} × unit time^{-1}) (output voxel grid size: 50×50 (unit length) ×20 (unit time), bandwidth of STKDE: h_1=100 (unit length), h_2=40 (unit time)). The color scale in these figures

is fixed in order to visually compare density difference among four scenarios. These maps show the summary of STP maps, which is useful to illustrate when and where particular movement behaviors were observed. In particular, these can highlight space-time bottlenecks explained by Cluster 5, 6, and 8 (Figure 128, Figure 129, and Figure 131) and successful evacuations explained by Cluster 3, 4, and 7 (Figure 126, Figure 127, and Figure 130).

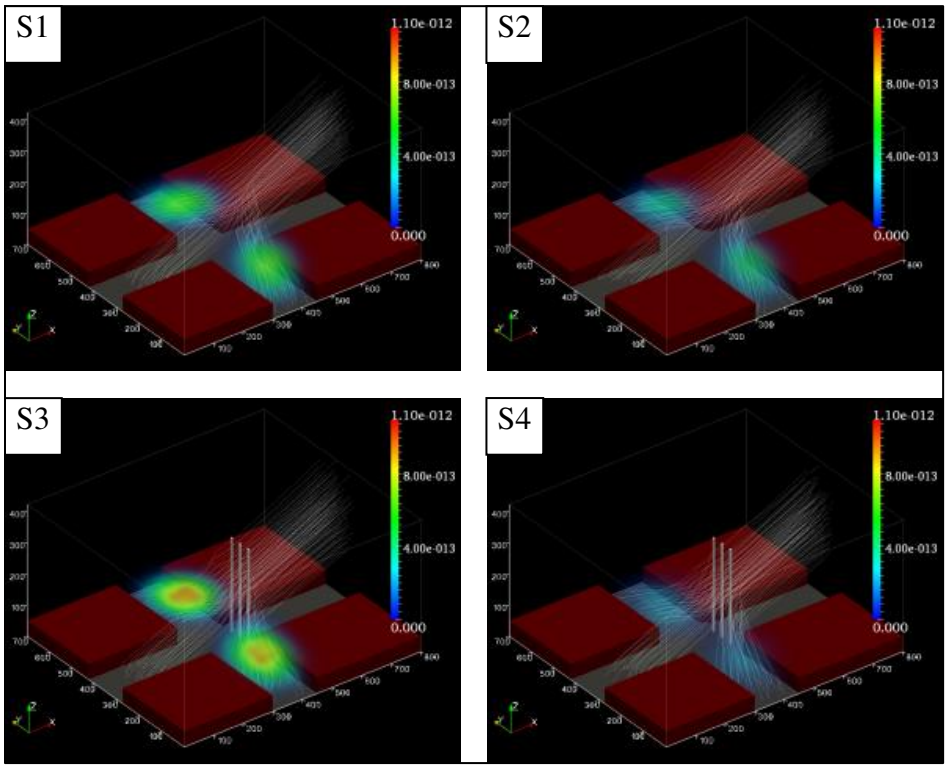

Figure 124. Space-Time line density map (Cluster 1: moderate velocity).

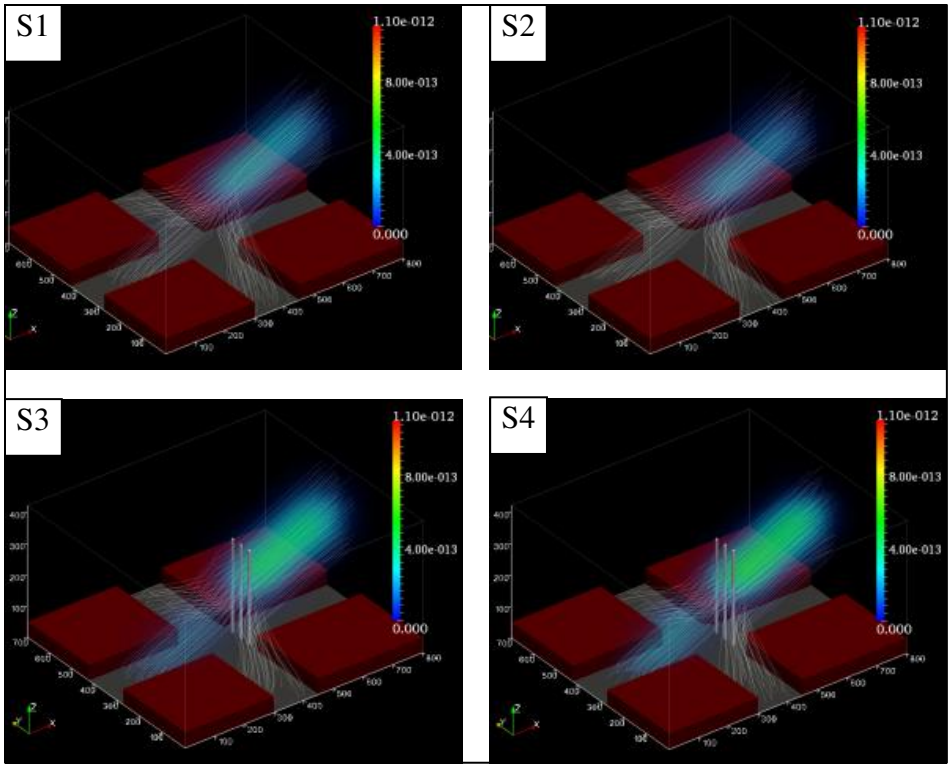

Figure 125. Space-Time line density map (Cluster 2: fragmented path).

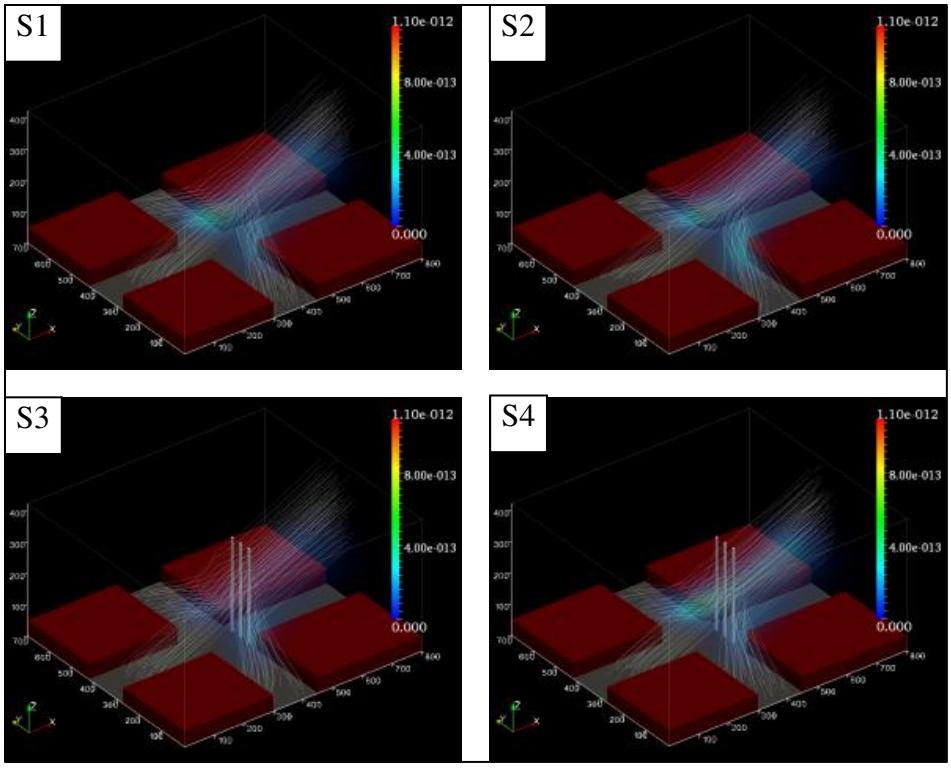

Figure 126. Space-Time line density map (Cluster 3: smooth & continuous).

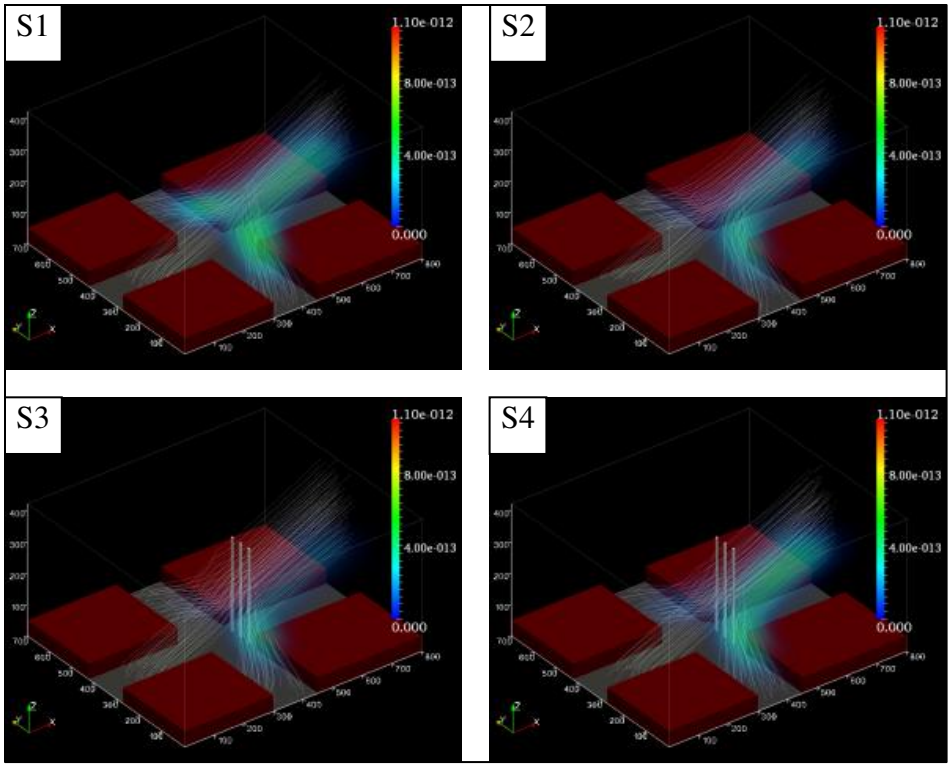

Figure 127. Space-Time line density map (Cluster 4: smooth & continuous).

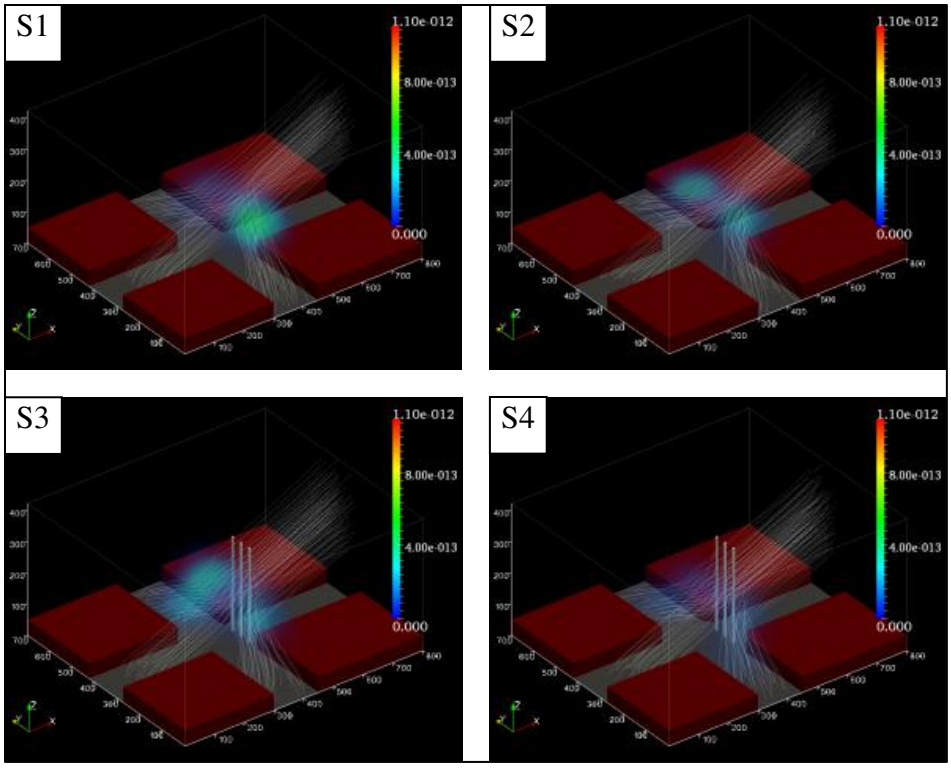

Figure 128. Space-Time line density map (Cluster 5: clogging).

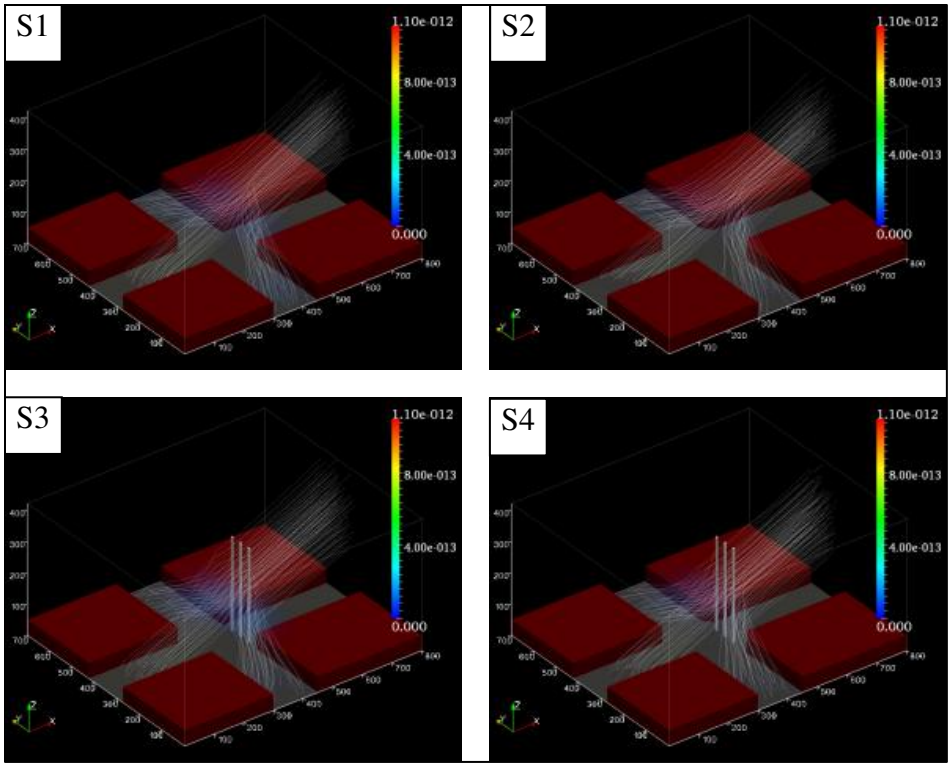

Figure 129. Space-Time line density map (Cluster 6: clogging).

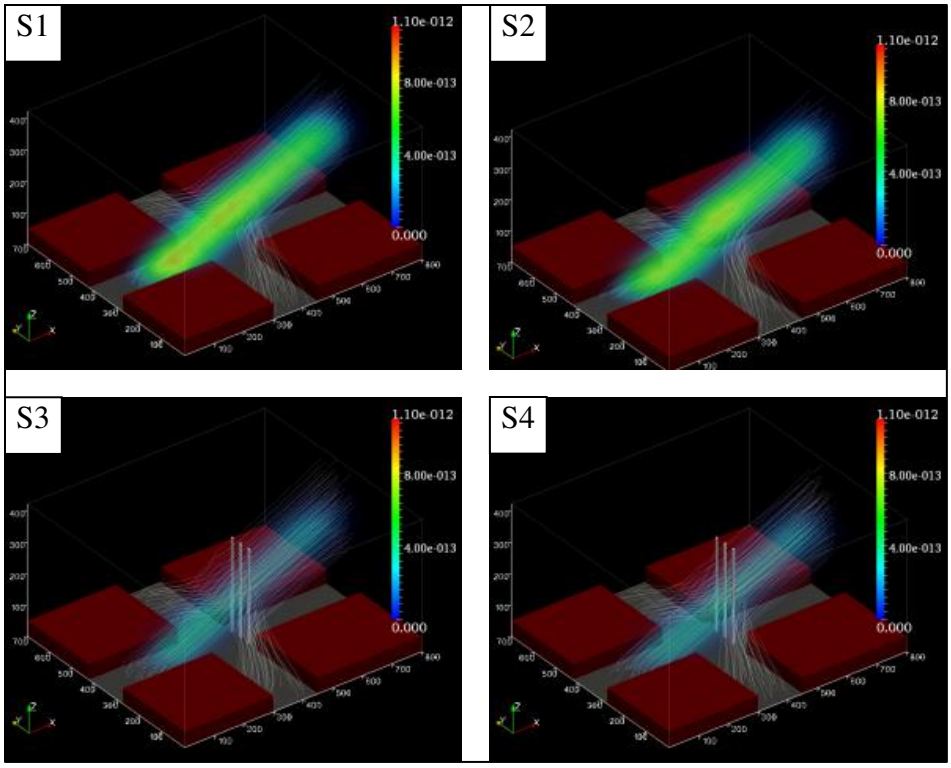

Figure 130. Space-Time line density map (Cluster 7: smooth & continuous).

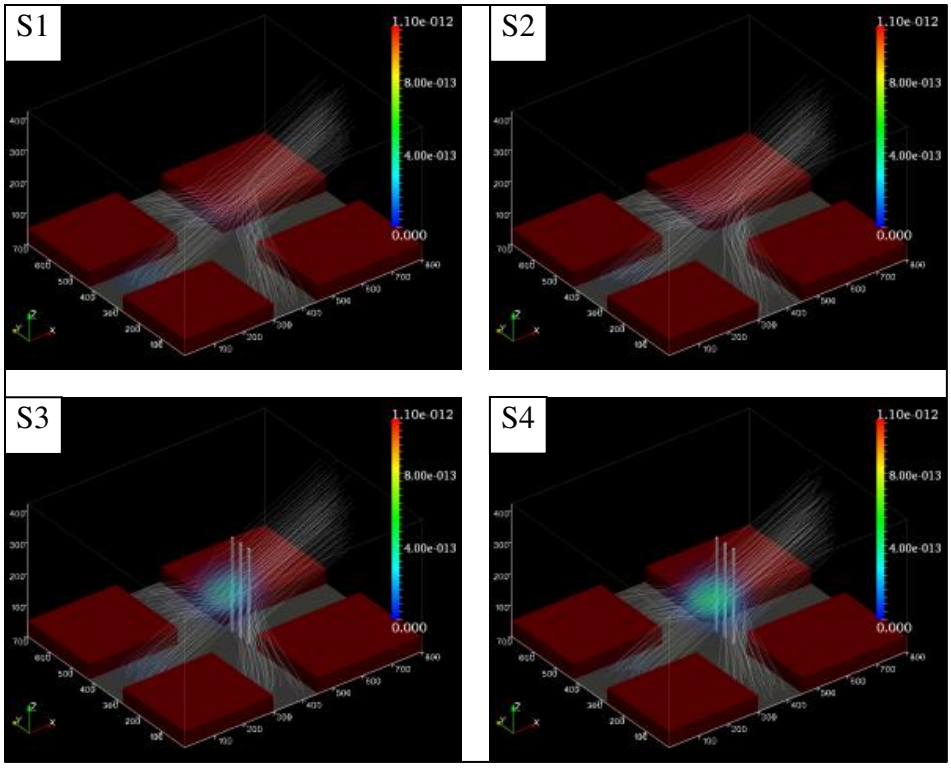

Figure 131. Space-Time line density map (Cluster 8: clogging).

In order to explicitly clarify the difference in spatio-temporal distribution of behavioral clusters, I applied 3D map algebra and utilized its local and subtraction operators to compare two space-time cubes. The approach is useful to spotlight the behavioral difference in space and time between two scenarios and helps in answering questions such as when and where a particular movement behavior is observed in one scenario and not in others, and how and why different movement behaviors appear in space and time. Figure 132 to Figure 139 illustrate the results of cluster density distribution differences between scenarios visualized in Space-Time Cubes for each cluster. The value of density difference describes the spatio-temporal intensity of crowd flow estimated by trajectory lengths (i.e., space-time line density). Because the intensity of density differences is varied in each cluster, I use different color scales for each figures. The center of the scale range is set to 0 and maximum (positive) and minimum (negative) value of the range is determined by the absolute value of density difference from four scenarios within a cluster. The value of density difference nearly equal to 0 is not displayed since the primary interest is the extreme difference in the density value. The label in each image denotes the subtraction operation between two scenarios. For example, “S1-S2” means that the density value of each voxel in Scenario 2 is subtracted from that in Scenario 1. In this case, positive values with warm color represent that Scenario 1 has higher values of the cluster density than Scenario 2, and negative values with cold color are vice-versa.

Between Scenario 1 (Base) and 2 (Rounded corners), significant difference is not displayed through all clusters. This result matches the results of

comparison by global and temporal approaches: that behavioral differences between Scenario 1 and 2 are much smaller than those between Scenario 1 and 3 or Scenario 1 and 4 (Figure 107 and Figure 113). However, identifying small differences can also help in understanding the effect of rounded corners; therefore, I created cluster density difference maps with different scale ranges for each cluster (Figure 140). The scale ranges are determined by the maximum and minimum values of the density difference between Scenario 1 and 2 within each cluster in order to exaggerate differences. The key behavioral difference visualized is Cluster 2 and 7, which have the largest and the second largest behavioral differences between two scenarios described by the comparison of global analysis (Figure 107). These two cluster distributions show two-layered flow, but the order of layer is inversely related. This describes several interesting crowd behaviors. First, the red flow in the bottom-left image in Figure 140 illustrates more successful evacuees from the West (Cluster 7) in Scenario 1 in the earlier stage of the simulation run. In fact, the crowd flow described by Cluster 7 is more concentrated in Scenario 1 than in Scenario 2 because the rounded corners created some spaces for evacuees from the West in Scenario 2. Second, the blue flow in the same image, which is on top of the red flow, describes more successful evacuees from the West in Scenario 2 in the later stage of the simulation. This explains that the successful flow (Cluster 7) in Scenario 2 persisted longer than in Scenario 1 because of, again, the space created by rounded corners. Third, the red flow in the top-right image in Figure 140 describes more unsuccessful evacuees (fragmented paths in Cluster 2) in the later

stage in Scenario 1, while the same behavior is observed in the earlier stage in Scenario 2 (blue flow), that is another effect of rounded corners. Even though these behavioral differences are interesting to show, again, the difference between Scenario 1 and 2 is subtle.

Between Scenario 1 (Base) and 3 (Bollards), several hot-cold spots of behavioral clusters are visualized. As compared to the base scenario, two higher density spots of clogging behavior were identified in Scenario 3 (Cluster 6 in Figure 137 & Cluster 8 in Figure 139). Cluster 6, as described before, is clogging behavior near the intersection created by evacuees from North and South when they decelerated to make turns, while Cluster 8 represents the clogging behavior at the middle of intersection due to the effect of obstacles and the congestion created by Cluster 6. The spatio-temporal sequence of these behaviors are well visualized in the image where the z value of the high density spot of Cluster 6 is lower than that of Cluster 8 in the Space Time Cube. In Scenario 3, the decrease of successful evacuees from the West (Cluster 7) and the increase of fragmented paths (Cluster 2) are also captured in Figure 133 and Figure 138 respectively. The reddish spot in Figure 138 illustrates the flow of successful evacuees from the West at the early stage in Scenario 1 is larger than that in Scenario 3. On the other hand, the blue spot in Figure 133 describes the flow of unsuccessful evacuees at the later stage in Scenario 3 is larger than that in Scenario 1.

Between Scenario 1 (Base) and 4 (Mixed), similar results by the comparison between Scenario 1 and 3 are identified, such as the decrease of the successful evacuees from the West (Cluster 7) and the increases of clogging

(Cluster 8) and fragmented paths (Cluster 2) in Scenario 4. These show the effect of inserting obstacles. In addition, the decrease of clogging described by Cluster 5 and no difference in clogging by Cluster 6 in Scenario 4 are observed, that is different from the result between Scenario 1 and 3. Because Cluster 5 and 6 exhibit clogging behaviors at the early stage of the simulation run, created by evacuees from North and South making turns, the effect of rounded corners reduced these behaviors and that effect are successfully visualized in the Space-Time Cube (Figure 136 and Figure 137).

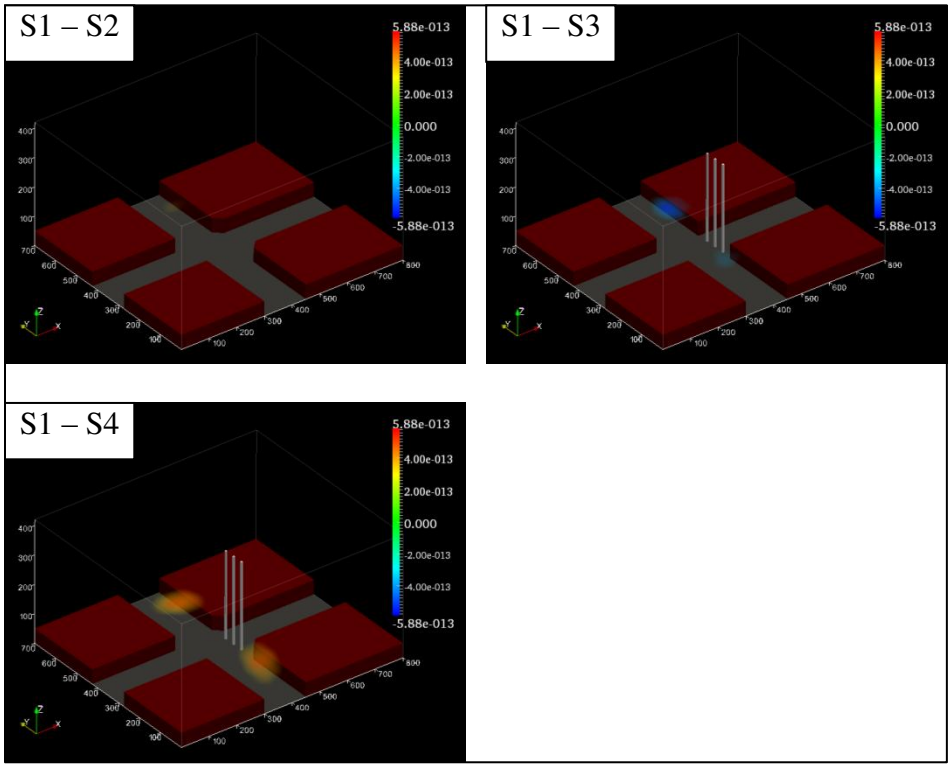

Figure 132. Difference of cluster density distribution between scenarios (Cluster 1).

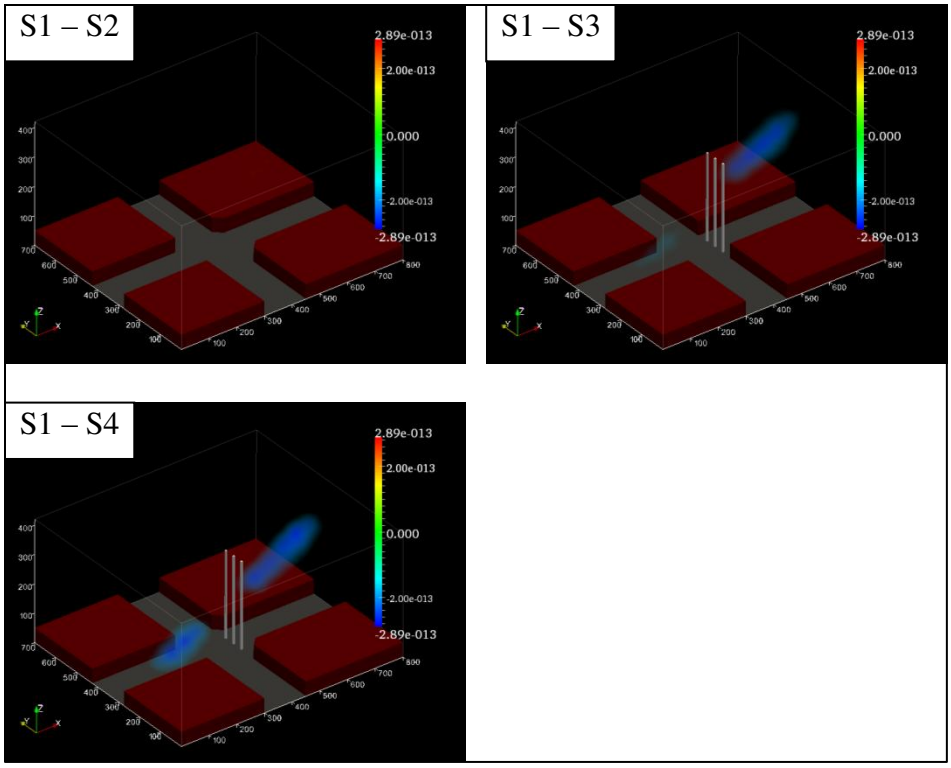

Figure 133. Difference of cluster density distribution between scenarios (Cluster 2).

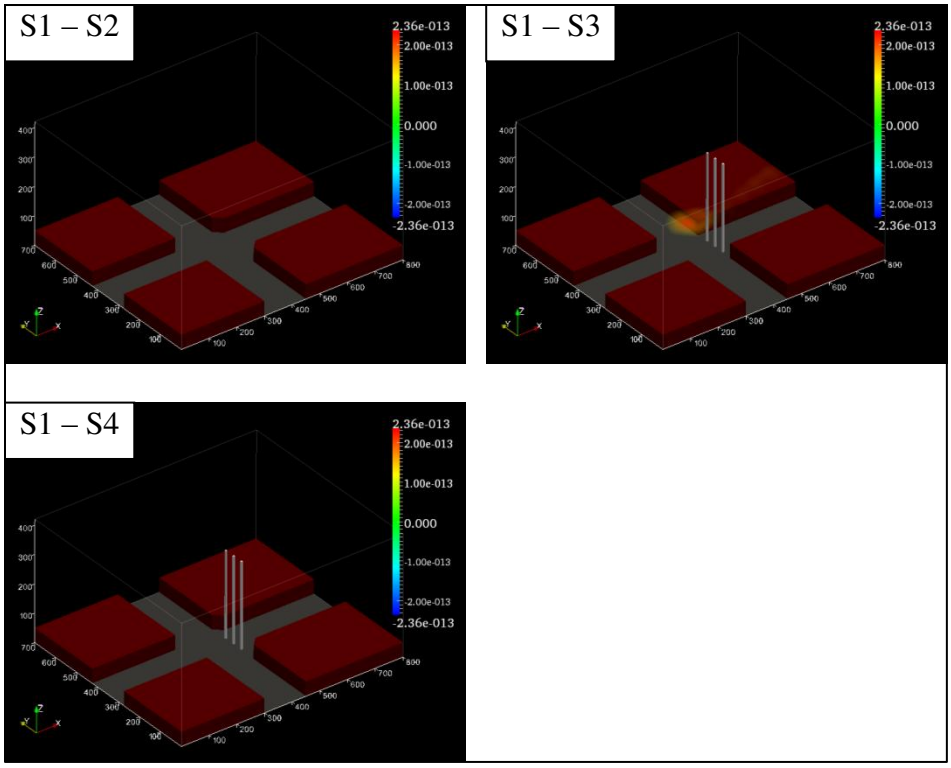

Figure 134. Difference of cluster density distribution between scenarios (Cluster 3).

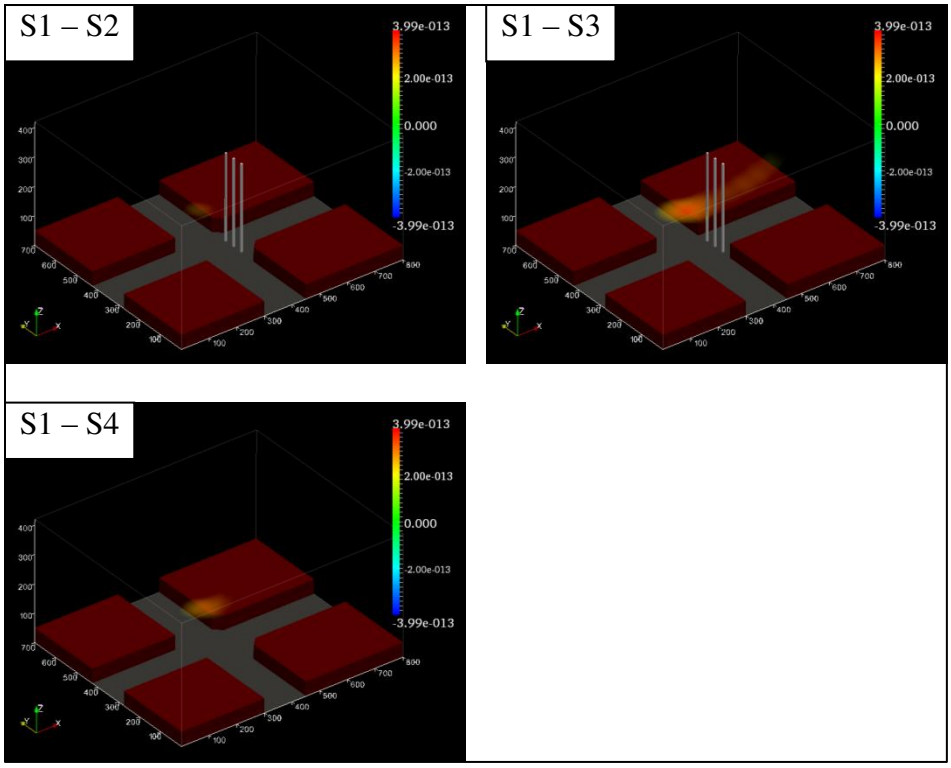

Figure 135. Difference of cluster density distribution between scenarios (Cluster 4).

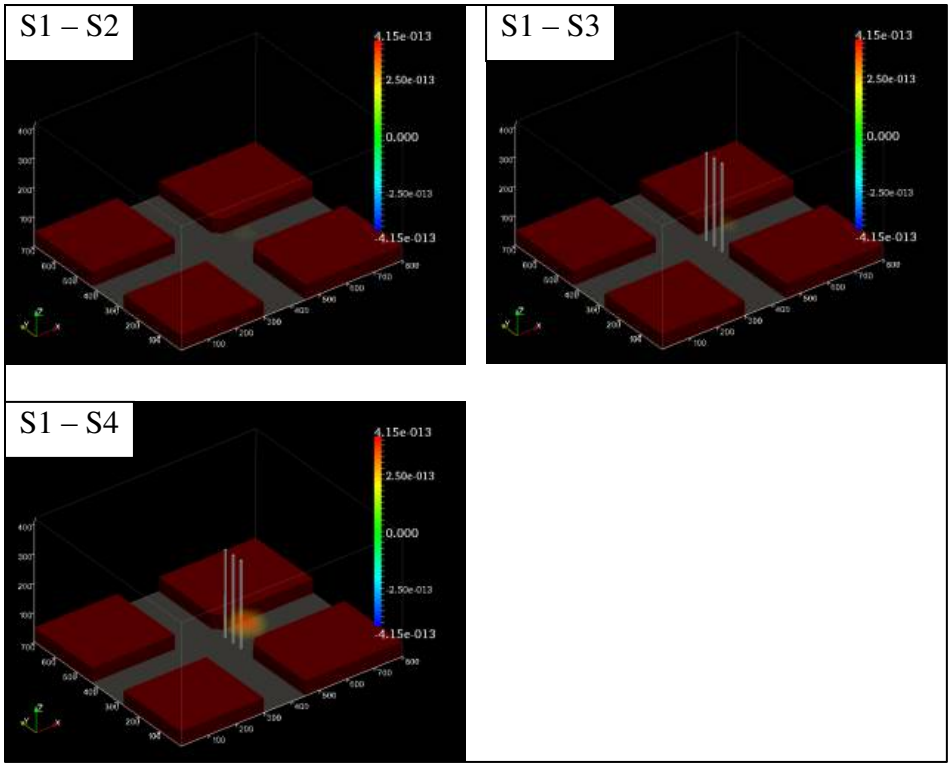

Figure 136. Difference of cluster density distribution between scenarios (Cluster 5).

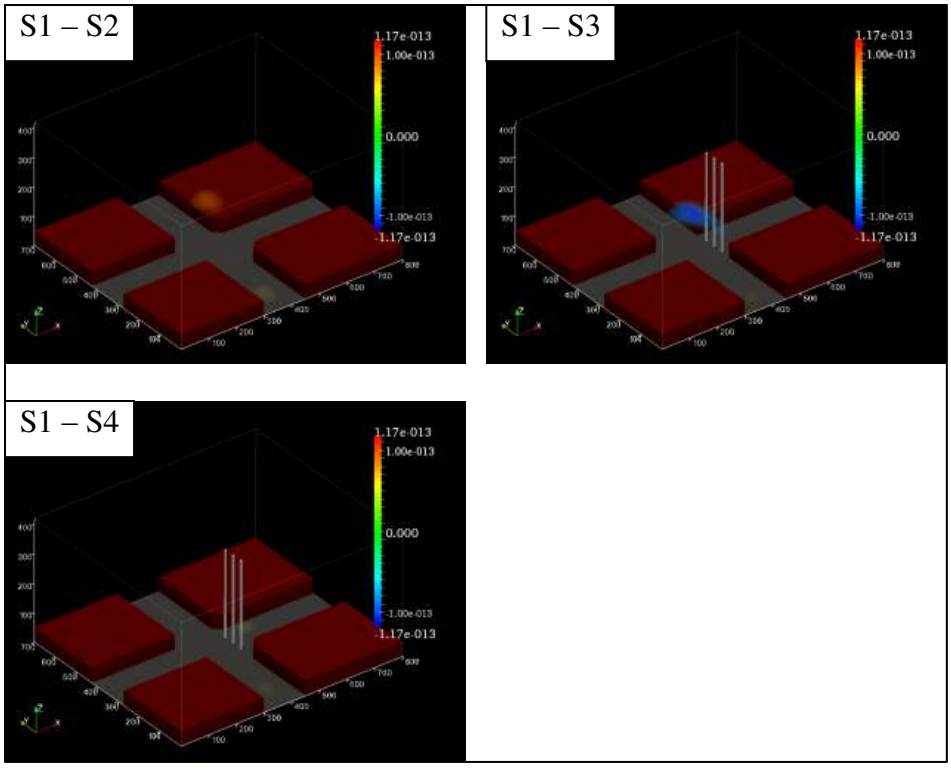

Figure 137. Difference of cluster density distribution between scenarios (Cluster 6).

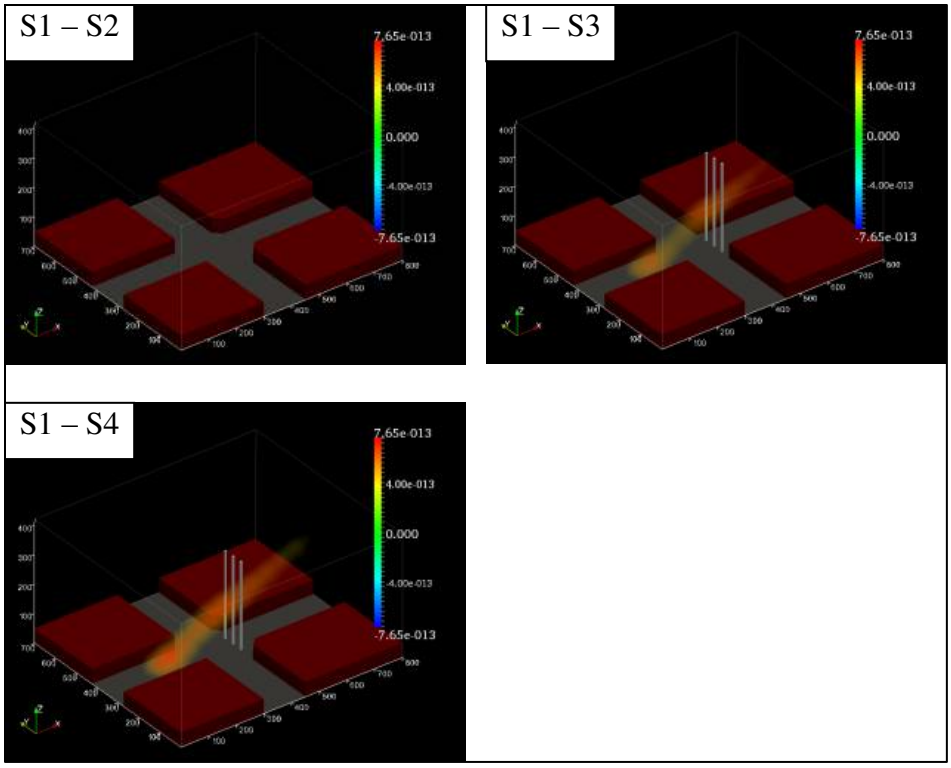

Figure 138. Difference of cluster density distribution between scenarios (Cluster 7).

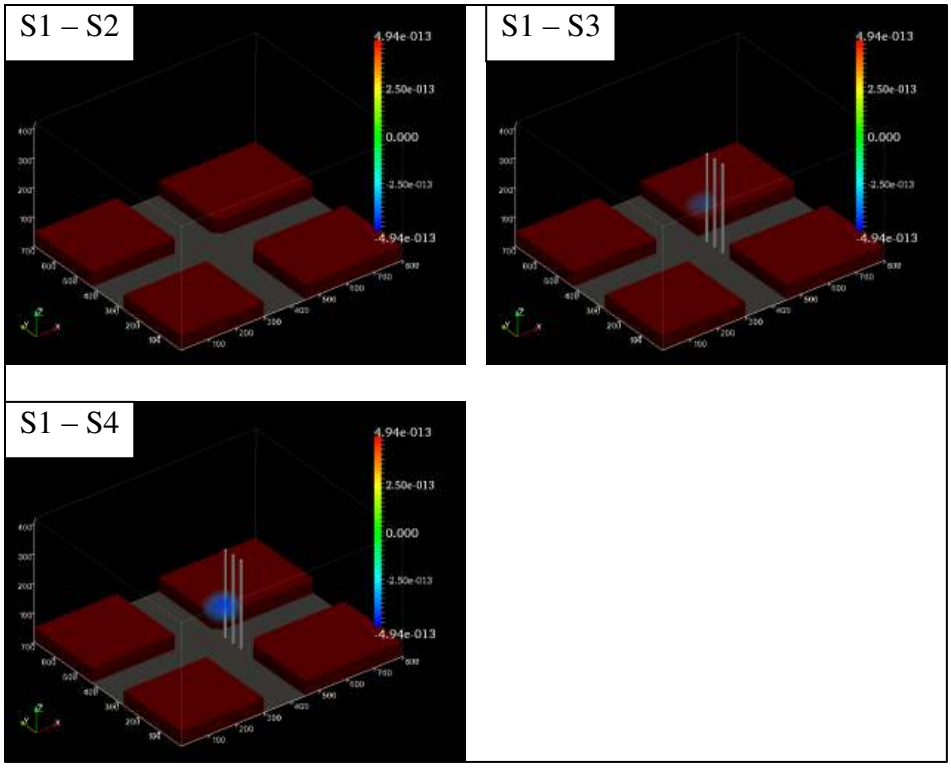

Figure 139. Difference of cluster density distribution between scenarios (Cluster 8).

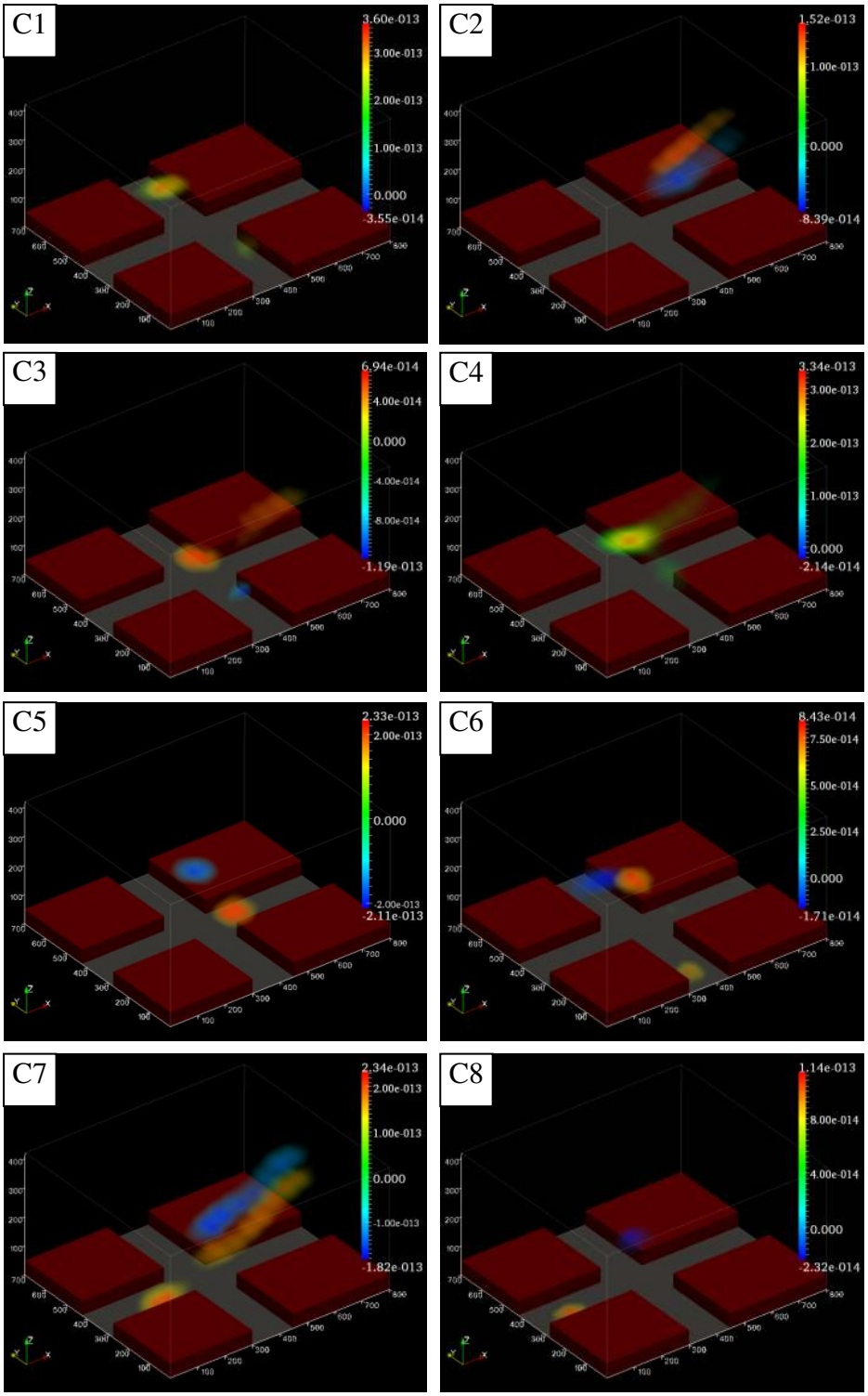

Figure 140. Detail difference of cluster density distribution between Scenario 1 and 2.

6.5 Discussion and Conclusion

This study was aimed at developing a new analytical framework for evaluating ABMs of mobile objects. I specifically focused on the research challenge of which aspects of the model behavior should be compared in order to capture complex movement behaviors and to validate simulation models. The proposed framework utilizes a trajectory data-mining approach to extract behavioral clusters from trajectory datasets determined by multiple motion characteristics. The extracted behavior clusters and their relationships in space and time can describe complex movement dynamics; thus, it is useful for model validation in finding spatial and temporal similarity and dissimilarity of behavioral clusters between the real-world and a model, between different models, or between different simulation scenarios. Three comparison methodologies for behavioral clusters were proposed; global, temporal, and spatio-temporal comparisons. The global approach compared the overall movement behaviors based on summarized behavioral cluster distribution, which helped in capturing general behavioral differences. The temporal approach compared the distribution of behavioral cluster through time, which enabled examination of when a particular behavior occurred, if there was a time lag of the occurrence of a particular behavior in different datasets, and if so why such a lag was observed. The spatio-temporal approach investigated and compared the distribution through space and time, which helped in answering questions such as when and where a particular movement behavior is observed in one dataset and not in others, and how and why such different movement behaviors appeared. The third approach specifically

employed the concept of time geography, in particular, STPs, ST-Volume density, and three-dimensional map algebra to capture and visualize differences of behavioral cluster distributions in space and time.

To examine the capability of the proposed framework for evaluating ABMs, I conducted an experiment to evaluate evacuation dynamics, which were generated by the social force model. The objective of the evaluation was to compare evacuation efficiency under four different designs of a four-way intersection. As a result, the proposed trajectory data-mining framework was found to be a useful approach for evaluation of crowd evacuation models by comparing behavioral clusters. Extracted behavioral clusters described collective movement behaviors during evacuation such as smooth and continuous movement, clogging, fragmented path, and zipper patterns due to clogging. The quantitative and qualitative comparison of behavioral clusters in four simulation scenarios enabled identification of similarity and dissimilarity among simulations, which successfully explained the effects of different designs of intersection, namely rounded corners and obstacles. The results showed that the impact of rounded corners improved evacuation efficiency by increasing the number of successful evacuees and decreasing the number of unsuccessful ones, whereas the influence of obstacles was vice-versa. This result is reasonable because rounding corners created spaces and encouraged evacuees to make smooth turns and the obstacles limited the available space for evacuees and created the bottlenecks.

The findings are perhaps obvious and already known; however, the approach to reach the conclusion is completely different from conventional

approaches such as global descriptive statistics (e.g., total egress time, average speed) and fundamental diagrams (e.g., flow-density, flow-speed, speed-density). Whereas conventional approaches consider basic properties of movement ignoring complex movement behaviors, the proposed approach takes multiple motion characteristics and collective behavior of mobile objects into account, which provides further insights into complex crowd dynamics. In particular, the result explained complex properties of evacuation dynamics such as how clogging behaviors were created at the early stage of evacuation and propagated through the congestion at the intersection when three flows merged together. In addition, it also showed that different intersection designs can amplify or curtail the size of congestion as a feedback process. Moreover, these findings are successfully mapped in Space-Time Cubes that allow visually capturing and exploration of such behaviors. Visual analyses combined with quantitative explanations described in this study are a very useful approach for evacuation management because decision-makers can visually identify hot/cold spots for evacuation efficiency. As interest in modeling mobile objects by incorporating complex behaviors grows, the proposed framework presents strong capability for evaluating ABMs of any kind of mobile objects by comparing complex behavioral movements between simulation and data from the real-world, different models, and different scenarios.

There is plenty of room for more experiments to advance the analytical power of the methodological framework (see 4.5 and 5.5). Particularly, it is important for the evaluation of ABMs to spatially and temporally investigate

multi-scaling effects. One solution is the use of STKDE with various voxel grid size and kernel bandwidth in order to summarize motion behaviors for investigating multi-scale movement effects. In addition, this study examined simulation of single run for four scenarios. In order to generalize the pattern and tendency in each scenario, it is necessary to examine multiple runs of each scenario. Because this study implemented three-dimensional map algebra, multiple runs of simulation can be summarized by using an averaging operator.

Regarding the development of ABM, the evacuation dynamics investigated in this study is a simple form of the social force model, where the characteristic of each pedestrian is not unique. Further exploration is required to better understand evacuation behavior by incorporating such as interrelationship of individuals (e.g., family, friends, and leader) and pedestrian behaviors such as panic, steering, and path-planning. Evaluating such models with complex behaviors using the developed trajectory data-mining framework is a goal of future work.

Chapter 7

SUMMARY

Recent advancements in Location Aware Technologies (LATs) allow researchers access to an unprecedented amount of data about individual mobile objects that until now were all but impossible. While collecting such data by LATs might be limited by cost, privacy, and security issues, Agent Based Models (ABMs) can realistically generate a massive collection of data about individual movements. These two sources of massive individual-scale movement data offer opportunities for investigating behavior of mobile objects in completely new ways. Specifically, extracting hidden patterns, trends, and useful information and knowledge from such massive and complex trajectory data is an emerging research area in Geographic Information Science (GIScience).

The research described here intends to contribute to the existing state-of-the-art in tracking and modeling mobile objects, in particular targeting challenges in extracting spatio-temporal patterns, processes, and useful knowledge from massive trajectory datasets. Specific research focuses are on the following challenges; 1) a lack of space-time analysis tools; 2) a lack of studies about empirical data analysis and context awareness (semantics) of movement datasets, particularly those considered as trajectories; and 3) a lack of studies about how to evaluate and test Agent-Based Models (ABMs) of mobile phenomena particularly focusing on a complex spatio-temporal and behavioral process of mobile agents.

To tackle these challenges, this dissertation conducted three studies on space-time analysis and modeling with following research objectives.

Study 1

- Developing an integrated spatio-temporal data exploration tool to represent spatio-temporal patterns and processes of mobile objects.
- Incorporating the framework of time geography for qualitative visualization of mobile objects.
- Incorporating quantitative representation of mobile objects.

Study 2

- Developing a trajectory data-mining methodology for context awareness of human movement.
- Generating theoretical movement data by random walk models.
- Collecting data of human spatio-temporal movements by GPS.
- Analyzing movement datasets with a spatio-temporal data exploration tool and trajectory data-mining methods.

Study 3

- Developing an agent-based simulation model of pedestrian evacuation dynamics to explore complex pedestrian behaviors.
- Quantitatively and qualitatively extracting pedestrian complex behaviors using the spatio-temporal data exploration tool and trajectory data-mining methods for evaluation of simulation models.

7.1 Achievements and Findings

The overarching goal of this research was to improve upon the current state-of-the-art in spatio-temporal analysis and modeling of complex human movement. To achieve this goal, I conducted three cohesive and interconnected studies on human trajectory data based around tool development, space-time analysis, visualization, data-mining, simulation, and model evaluation. In summary, the first study discussed development of a toolkit that could quantify trajectory datasets and qualitatively visualize the quantitative results within the scope of time geography. The second study extended the toolkit by implementing the trajectory data-mining tool to further investigate trajectory datasets, and the third study applied the toolkit to evaluation of an ABM of crowd evacuation dynamics. The following sections present detail achievements with respect to the research objectives in each study.

7.1.1 Study 1

In the first study, a novel spatio-temporal data exploration toolkit was developed to analyze and represent spatio-temporal patterns and processes of mobile objects. The toolkit integrated both quantitative and qualitative representations of mobile objects. As a quantitative representation, the toolkit calculates various motion descriptors to characterize individual trajectories including basic motion descriptors (i.e., velocity, acceleration, orientation, length, and sinuosity), fractal dimension, directional distribution, and Lévy metrics.

As a qualitative representation, the toolkit implemented a visualization technique based around the concept of time geography. Specifically, a trajectory dataset can be visualized in a Space-Time Cube as Space-Time Paths (STPs), which can be enhanced by the use of color and tube representations based on calculated scalar values of motion descriptors. In addition, the toolkit allows estimating Space-Time volume density of trajectory datasets by Space-Time Kernel Density Estimation (STKDE), which ultimately produces Space-Time volume density maps of trajectory datasets. These quantitative and qualitative representations provide new insights for understandings spatio-temporal behavioral patterns and processes in large and complex data of mobile objects.

The case study demonstrates that collective movement behaviors of pedestrian crowds under evacuation scenarios can be described, even for massive amount of data and for complex scenarios with many interacting movements. The results capture and describe collective behavior of crowd congestion, an important feature of evacuation dynamics, in detail in space and time. Such results can be used for better facility design as well as decision-makings about evacuation route planning and scheduling.

In addition, the toolkit provides a Graphic User Interface (GUI) for efficiency and ease of use for various tools implemented in the toolkit including data manipulation tools as well as analytical tools.

7.1.2 Study 2

The second study seeks to enhance the capability of the developed toolkit in the first study to further investigate movement behaviors of mobile objects specifically focusing on behavioral context recognition. The goal is achieved through integrating the trajectory data-mining function with the developed tool. The function includes trajectory partitioning and clustering algorithms to extract behavioral patterns of mobile objects using multiple motion descriptors as well as visual analysis to display extracted patterns and trends in space and time. The extracted behavioral clusters are further used for behavioral recognition of mobile objects.

Two case studies were performed to examine the functionality. The first case study examined the dataset generated by pure mathematical models so that their movement behaviors are known. Therefore, it is useful to examine how well the trajectory data-mining function performs. The dataset consists of mixed trajectories simulated by three random walk models; Brownian Motion (BM), Correlated Random Walk (CRW), and Lévy flight. The second case study examined real-world trajectory dataset, which were collected by a GPS device.

The results demonstrated that local behaviors of trajectory were well extracted and they were able to explain the global behavioral context from mixed trajectories of random walkers. Extracted local behaviors in the GPS dataset differentiated real movement activities during a day; however, the explanation power for global behavioral context recognition by local behaviors was not much improved from the recognition by global behaviors. These results indicate that the

proposed trajectory data-mining framework performs well on mixed behavioral datasets that are explicitly defined by mathematical expressions; however, when it applied to the real-world dataset to understand complex behaviors of human movements, the explanation power is limited.

7.1.3 Study 3

The third study applied the toolkit developed in study 1 and 2 to evaluate an ABM of crowd dynamics under evacuation. Specifically, the study proposed to use the trajectory data-mining toolkit for model validation by extracting behavioral clusters of collective movements from simulation, and to compare the distribution of the extracted clusters against a dataset from the real-world, an other simulation model, or different scenarios. Three comparison methodologies for the distribution of behavioral clusters were proposed; global comparison, temporal comparison, and spatio-temporal comparison. The spatio-temporal approach, in particular, investigates and compares the distribution through space and time by employing the time geography framework of STPs, space time volume density, and three-dimensional map algebra. This allows capturing and visualizing differences of behavioral cluster distributions in space and time in different models.

To examine the capability of the proposed framework for evaluating ABMs, I conducted an experiment to evaluate evacuation dynamics at a four-way intersection. The trajectory data were generated by the social force model. The

objective of the evaluation was to compare evacuation efficiency under four different designs of a four-way intersection.

The results demonstrated that the trajectory data-mining framework is a useful approach for evaluation of crowd evacuation models. Extracted behavioral clusters described collective movement behaviors during evacuation such as smooth and continuous movement, clogging, fragmented path, and zipper patterns due to clogging. Quantitative and qualitative comparison of behavioral clusters in four simulation scenarios enabled identification of behavioral similarity and dissimilarity among simulations, which successfully explained the effects of different designs of intersection, namely rounded corners and obstacles. The results showed that the impact of rounded corners improved the evacuation efficiency by increasing the number of successful evacuees and decreasing the number of unsuccessful ones, whereas the influence of obstacles was vice-versa. This result is reasonable because rounding corners created spaces and encouraged evacuees to make smooth turns and obstacles limited the available space for evacuees and created the bottlenecks.

7.2 Limitations

There are some limitations in this study. First, even though the developed trajectory data mining framework can deal with trajectory dataset with multiple mobile objects (see 0), behavioral context recognition of real-world trajectories (0) was based on a single person's daily GPS data (n=36) due to the limited data availability. By using dataset of multiple mobile objects, the framework could

map clusters of movement behavior to capture spatio-temporal pattern and tendency. For example, a commuter town is likely to show a high density cluster distribution of commuting behavior (e.g., relatively fast and directed movement) in morning and evening, whereas a shopping district is likely to show a high density cluster distribution of shopping behavior (e.g., relatively slow and wondering movement).

Second, the social force model in Study 1 (0) and 3 (0) generated trajectories with simple evacuation behavior, which limits behavioral complexity. Even though variability was introduced as a parameter of agent's desired velocity by a probability function using Gaussian distribution, agent's movement behavior was homogeneously modeled by the simple social force model. This might be realistic in some emergency situations where people perceive risk from the response of others and behave similarly each other; however, in many situations, behavior can be heterogeneous and far more complex. For example, social relationships (e.g., family, friend) may create flocking behavior, social roles (e.g., superior and subordinate) may create leading and following behavior, physical ability (e.g., age, disability) may introduce various movement behaviors in terms of such as walking speed, vision, and accessibility, and personal characteristics and psychological effects may lead panic behavior. Incorporating these behaviors into an ABM would produce more rich, complex, and realistic movement behaviors. Evaluation of the developed trajectory data-mining framework will be better achieved by extracting such complex behaviors.

Third, in Study 3 (0), I focused on the evacuation dynamics on a four-way intersection with a unidirectional flow. Although the results successfully extracted some behavioral complexities and examined the evacuation performance under four different street designs of intersection, more exploration is required to fully examine the capability of the developed framework. For example, simulation can be run under different infrastructural designs such as building with multiple floors, multiple exists, and multiple flows.

Forth, the size of a voxel grid and a kernel bandwidth for estimating space-time volume density were empirically defined and fixed at one scale in all three studies. This limits the capability to capture multi-scaling effects spatially and temporally of movement behavior, which are typically found in a complex system. One solution is to use various voxel grid size and kernel bandwidth to summarize motion behaviors for investigating multi-scale movement effects such as goal-oriented movement at macro-scale (e.g., work to home) and wandering movement at micro-scale (e.g., shopping on the way to home, wandering of pedestrian on the street due to high crowd density).

7.3 Discussions and Future Works

This research aims to investigate human spatio-temporal behaviors in three ways. The first study develops a spatio-temporal data exploration tool, which enables us to qualitatively and quantitatively investigate spatio-temporal patterns of mobile objects. The second study explores simulated and empirical human trajectory datasets to understand movement activities in space and time by retrieving

behavioral contexts using the spatio-temporal data exploration tool and trajectory data-mining method. Finally, the third study investigates the behavioral process of pedestrian collective movement by developing an agent-based simulation model and analyzing simulation outcomes with the spatio-temporal data exploration tool and trajectory data-mining method.

The potential impacts of this study are broad. The research contributes to understanding of human dynamics inductively and deductively. The second study uses the toolkit inductively to exploratory analyze mobile objects. This helps understanding of complex human-environment interaction and thus formulating hypotheses in behavioral geography such as spatial cognition, decision-making and choice behaviors in mobility, and collective movement. The third study is a deductive approach in which a pedestrian simulation model is developed based on existing theories and the toolkit is used for the model evaluation exercise. The scientific approach of these inductive-deductive loops gives further insight into the study of individual-scale human movement, focusing on its behavioral patterns and processes. Methodologically, this research develops a novel analytical tool to investigate spatio-temporal behaviors at various spatio-temporal scales from street to city and from second to day respectively. Potential practical implications are numerous such as decision-making and decision support systems for urban planning, facility design, and socio-behavioral planning. Specifically, such applications include vehicle and pedestrian traffic control for transportation and pedestrian facilities design and management (e.g., congestion management, crowd control, and evacuation), location-based services (e.g., navigation and

advertisement); and law enforcement (e.g., video surveillance for criminal activities).

There are several considerations for future work. First, more experiments are required to advance the analytical power of the methodology and toolkit; for example, fine-tuning of model parameters particularly concerning spatial and temporal granularity (e.g., resampling frequency, parameters for trajectory partition algorithms, *k* value in k-means clustering, grid size and band width selection for STKDE), variable selection of motion descriptors, methodological exploration with other motion characterization (e.g., incorporating variances in addition to mean values), clustering, and classification techniques, and experiments with other dataset.

Second, the second study demonstrated limitations when applying the toolkit to real-world trajectory data. When recognizing the behavioral activities of a trajectory dataset, this study only considered the composition of local behaviors extracted by trajectory data-mining. One potential solution to improve the inference of complex activities is to use additional information such as locational information and temporal sequence of trajectory clusters instead of just using the composition of trajectory clusters.

Third, regarding agent-based modeling, this study examined simulations of a single run for four scenarios. In order to generalize the pattern and tendency in each scenario, it is necessary to examine multiple runs of each scenario. Because this study implemented three-dimensional map algebra, multiple runs of simulation can be summarized by using the averaging operator. Furthermore,

evacuation dynamics investigated in this study represent a simple form of the social force model, where the characteristic of each pedestrian is not unique. Further exploration is required to better understand evacuation behavior by incorporating considerations such as interrelationship of individuals (e.g., family, friends, and leader) and pedestrian behaviors such as panic, steering, and path-planning. Evaluating such models with complex behaviors using the developed trajectory data-mining framework is a topic for future work.

REFERENCES

Abdul-Rahman, A., & Pilouk, M. (2008). *Spatial Data Modeling for 3D GIS.* Berlin: Springer-Verlag.

Active Bat. (2009). *Active Bat.* Retrieved 12 30, 2009, from http://www.cl.cam.ac.uk/research/dtg/attarchive/bat/

Adler, T. J., & Ben-Akiva, M. E. (1979). A theoretical and empirical model of trip chaining behavior. *Transportation Research B, 13*, 243-257.

AEA Technology. (2002). *A Technical Summary of the AEA EGRESS Code.* Technical Report, AEAT/NOIL/27812001/002(R), Issue 1.

Agrawal, R., Gehrke, J., Gunopulos, D., & Raghavan, P. (1998). Automatic Subspace Clustering of High Dimensional Data for Data Mining Applications. *International Conference Management of Data (SIGMOD)*, (pp. 94-105).

Alberti, M., & Waddell, P. (2000). An integrated urban development and ecological simulation model. *Integrated Assessment 1*, 215-227.

Allen, P. A. (1981). The evolutionary paradigm of dissipative structures. In E. Jantsch (Ed.), *The Evolutionary Vision* (pp. 25-71). Boulder: Westview Press.

Alonso, W. (1960). A theory of the urban land market. *Papers and Proceedings of the Regional Science Association*(6), 149-158.

Andrienko, G., & Andrienko, N. (1999). Data mining with C4.5 and interactive cartographic visualization. *User interfaces to data intensive systems*, (pp. 162-165). Los Alamitos, CA.

Andrienko, G., Andrienko, N., Rinzivillo, S., Nanni, M., Pedreschi, D., & Giannotti, F. (2009). Interactive Visual Clustering of Large Collections of Trajectories. *Proceedings of IEEE Symposium on Visual Analytics and Technology*, (pp. 3-10).

Andrienko, N., & Andrienko, G. (2011). Spatial generalization and aggregation of massive movement data. *IEEE Transactions on Visualization and Computer Graphics, 17*(2), 205-219.

Andrienko, N., Andrienko, G., & Gatalsky, P. (2003). Visual data exploration using space-time cube. *Proceedings of the 21st International Cartographic Conference (ICC)*, (pp. 10-16). Durban, South Africa.

Ankerst, M., Breuning, M. M., Kriegel, H. P., & Sander, J. (1999). OPTICS: Ordering Points to Identify the Clustering Structure. *Proceedings of the ACM*

SIGMOD International Conference on Management of Data, (pp. 49-60). Philadelphia, PA.

Axhausen, K. W., & Gärling, T. (1992). Activity-based approaches to travel analysis: conceptual frameworks, models, and research problems. *Transport Reviews, 12*(4), 323-341.

Aydinonat, N. E. (2005). An interview with Thomas C. Schelling: Interpretation of game theory and the checkboard model. *Economics Bulletin, 2*(2), 1-7.

Bahl, P., & Padmanabhan, V. (2000). RADAR: An in-building RFbased user location and tracking system. *Proceedings of IEEE INFOCOM*, *2*, pp. 775-784.

Bailey, T., & Gatrell, T. (1995). *Interactive spatial data analysis.* Prentice Hall.

Bartumeus, F., da Luz, M. G., Viswanathan, G. M., & Catalan, J. (2005). Animal search strategies: a quantitative random-walk analysis. *Ecology, 86*, 3078-3087.

Bateman, J. M., & Edwards, B. (2002). Gender and evacuation: A closer look at why women are more likely to evacuate for hurricanes. *Natural Hazards Review*, 107-117.

Batschelet, E. (1981). *Circular Statistics in Biology.* London, UK: Academic Press.

Batty, M. (2003). Agent-Based Pedestrian Models. In P. A. Longley, & M. Batty, *Advanced Spatial Analysis: The CASA Book of GIS* (pp. 81-106). Redlands, CA: ESRI Press.

Batty, M. (2005). *Cities and Complexity.* Cambridge, UK: MIT Press.

Batty, M., & Longley, P. (1994). *Fractal Cities.* Academic Press.

Batty, M., & Torrens, P. M. (2005). Modeling and prediction in a complex world. *Futures, 37*, 745-766.

Batty, M., Fotheringham, S. A., & Longley, P. (1993). Fractal Geometry and Urban Morphology. In N. S. Lam, & L. D. Cola (Eds.), *Fractal in Geography* (pp. 228-246).

Ben-Akiva, M., & Bierlaire, M. (2003). Discrete choice models with applications to departure time and route choice. *International Series in Operations Research & Management Science, 56*(2), 7-37.

Benenson, I., & Torrens, P. M. (2004). *Geosimulation: Automata-Based Modeling of Urban Phenomena.* London, UK: John Wiley & Sons.

Benhamou, S. (2004). How to reliably estimate the tortuosity of an animal's path: straightness, sinuosity, or fractal dimension? *Journal of Theoretical Biology, 229*, 209-220.

Bergman, C. M., Schaefer, J. A., & Luttich, S. (2000). Cariboumovement as a correlated random walk. *Oecologia, 123*(3), 364–374.

Berrow, J. L., Beecham, J., Quaglia, P., Kagarlis, M. A., & Gerodimos, A. (2005). Calibration and validation of the Legion simulation model using empirical data. *Pedestrian and evacuation dynamics 2005*, 167-181.

Bertalanffy, L. V. (1968). *General System Theory.* New York: George Braziller.

Bishop, C. M. (1995). *Neural Networks for Pattern Recognition.* Oxford, UK: Oxford University Press.

Blue, V. J., & Adler, J. L. (1999). Using cellular automata. Microsimulation to model pedestrian movements. *Proceedings of the 14th International Symposium on Transportation and Traffic Theory*, (pp. 235-254).

Blue, V. J., & Adler, J. L. (2000). Cellular automata microsimulation of bidirectional pedestrian flows. *Transportation Research Board, 1678*, 135-141.

Bogorny, V., Kuijpers, B., & Alvares, L. O. (2009). ST-DMQL: A semantic trajectory data mining query language. *International Journal of Geographic Information Science, 23*(10), 1245-1276.

Borgers, A., & Timmermans, H. (1986). A model of pedestrian route choice and demand for retail facilities within inner-city shopping areas. *Geographical Analysis, 18*(2), 115-128.

Boschetti, F., Dentith, M. D., & List, R. D. (1996). A fractal-based algorithm for detecting first arrivals on seismic traces. *Geophysics, 61*(4), 1095-1102.

Bovy, P. H., & Stern, E. (1990). *Route Choice: Wayfinding in Transport Networks.* Dordrecht: Kluwer Academic Publishers.

Bowman, J. L., & Ben-Akiva, M. E. (2000). Activity-based disaggregate travel demand model system with activity schedules. *Transportation Research Part A, 35*, 1-28.

Brail, R. K., & Klosterman, R. E. (2001). *Planning Support System: Integrating Geographic Information Systems, Models and Visualization Tools.* Redlands, CA: ESRI.

Braun, A., Musse, S. R., de Oliveira, L. P., & Bodmann, B. E. (2003). Modeling individual behaviors in crowd simulation. *Proceedings of the 16th International*

Conference on Computer Animation and Social Agents (pp. 143-148). Washington, DC: IEEE Computer Society.

Brockmann, D., Hufnagel, L., & Geisel, T. (2006). The scaling laws of human travel. *Nature, 439*, 462-465.

Brunsdon, C., Corcoran, J., & Higgs, G. (2007). Visualising space and time in crime patterns: A comparison of methods. *Computers, Environment, and Urban Systems, 31*, 52-75.

Bryan, J. L. (1982). MGM grand hotel fire: a case study of human reaction to fire. *Proceedings of the 6th joint panel meeting of the UJNR*, (pp. 2185-250). Tokyo, Japan.

Bryan, J. L. (1995). Behavioural Response to Fire and Smoke. In *SFPE Handbook of fire protection engineering* (pp. 3-241-3-262). Quincy, MA: Society of Fire Protection Engineers.

Bryan, J. L. (1995). Behavioural Response to Fire and Smoke. In *SFPE Handbook of fire protection engineering* (pp. 3-241-3-262). Quincy, MA: Society of Fire Protection Engineers.

Burnett, P., & Hanson, S. (1982). The analysis of travel as an example of complex human behaviour in spatially-constrained situations: Definition and measurement issues. *Transportation Research A, 16*, 87-102.

Calabrese, F., Reades, J., & Ratti, C. (2010). Eigenplaces: Segmenting space through digital signatures. *IEEE Computer Society, 9*(1), 78-84.

Calenge, C. (2006). The package adehabitat for the R software: a tool for the analysis of space and habitat use by animals. *Ecological Modelling, 197*, 516-519.

Cao, H., Mamoulis, N., & Cheung, D. W. (2009). Periodic pattern discovery from trajectories of moving objects. In H. Miller, & J. Han (Eds.), *Geographic data mining and knowledge discovery* (pp. 389-408). CRC Press, Taylor and Francis Group.

Chainey, S., & Ratcliffe, J. (2005). *GIS and Crime Mapping.* John Wiley & Sons.

Chen, J., Shaw, S.-L., Yu, H., Lu, F., Chai, Y., & Jia, Q. (2011). Exploratory data analysis of activity diary data: a space-time GIS approach. *Journal of Transport Geography*.

Ciolek, T. M. (1981). Pedestrian behaviour in pedestrian spaces: some findings of a naturalistic field study. *Proceedings of the annual conference of the ANZAScA* (pp. 95-112). Canberra: Australian and New Zealand Architectural Science Association.

Clark, W. A., & Dieleman, F. M. (1996). *Households and housing: Choice and outcomes in the housing market.* New Brunswick, NJ: The Center For Urban Policy Research.

Clark, W. A., & Huang, Y. (2003). The life course and residential mobility in British Housing markets. *Environment and Planning A, 35*, 323-339.

Clark, W. A., Huang, Y., & Withers, S. (2003). Does commuting distance matter? Commuting tolerance and residential change. *Regional Science and Urban Economics, 33*, 199-221.

Corcoran, J., Higgs, G., Brunsdon, C., & Ware, A. (2007). The use of Comaps to Explore the spatial and temporal dynamics of fire incidents: A case study in South Wales, United Kingdom. *The Professional Geographer, 59*(4), 521-536.

Cornwell, B. (2003). Bonded fatalities: Relational and ecological dimensions of a fire evacuation. *Sociological Quarterly, 44*(4), 617-638.

Cortes, C., & Vapnik, V. (1995). Support-vector networks. *Machine Learning, 20*(3), 273-297.

Couch, C. J. (1968). Collective behavior: An examination of some stereotypes. *Social Problems, 15*, 310-322.

Crooks, A., Castle, C., & Batty, M. (2008). Key Challenges in Agent-Based Modeling for Geo-Spatial Simulation. *Computers, Environment, and Urban Systems, 32*(6), 417-430.

Daamen, W., & Hoogendoorn, S. P. (2003). Experimental research of pedestrian walking behavior. *Transportation Research Record, 1828*, 20-30.

Daamen, W., & Hoogendoorn, S. P. (2003). Qualitative results from pedestrian laboratory experiments. In E. R. Galea, *Pedestrian and evacuation dynamics* (pp. 121-132). London: CMS Press.

Damm, D., & Lerman, S. R. (1981). A theory of activity scheduling behavior. *Environment and Planning A, 13*, 703-718.

De la Barra, T. (1989). *Integrated Land Use and Transportation Modelling: Decision Chains and Hierarchies.* Cambridge: Cambridge University Press.

DeMers, M. N. (2002). *GIS modeling in raster.* Chichester, UK: John Wiley and Sons.

Dempster, A., Laird, N., & Rubin, D. (1977). Maximum likelihood from incomplete data via the EM algorithm. *Journal of Royal Statistical Society, 39*, 1-38.

Desyllas, J., Duxbury, E., Ward, J., & Smith, A. (2003). Pedestrian demand modelling of large cities: An applied example from London. *Centre for Advanced Spatial Analysis, Working Paper Series, Paper 62*.

Digman, J. M. (1990). Personality structure: Emergence of the five-factor model. *Annual Review of Psychology, 41*, 417-440.

Dodge, S., Weibel, R., & Forootan, E. (2009). Revealing the physics of movement: comparing the similarity of movement characteristics of diff erent types of moving objects. *Computers, Environment and Urban Systems, 33*(6), 419-434.

Dodge, S., Weibel, R., & Lautenschütz, A.-K. (2008). Toward a taxonomy of movement patterns. *information Visualization*.

Domingos, P., & Pazzani, M. (1997). On thee optimality of the simple Bayesian classifier under zero-one loss. *Machine Learning, 29*, 103-130.

Durupinar, F., Pelechano, N., Allbeck, J. M., Güdükbay, U., & Badler, N. I. (2011). The impact of the OCEAN personality model on perception of crowds. *IEEE Computer Graphics and Applications*, in press.

Dyer, J. R., Johansson, A., Helbing, D., Couzin, I. D., & Krause, J. (2009). Leadership, consensus decision making and collective behaviour in humans. *Philosophical Transactions of The Royal Society B, 364*(1518), 781-789.

Eagle, N., & Pentland, A. (2006). Reality mining: Sensing complex social systems. *Personal and Ubiquitous Computing, 10*(4), 255-268.

Eagle, N., & Pentland, A. (2009). Eigenbehaviors: Identifying structure in routine. *Behavioral Ecology and Sociobiology, 63*(7), 1057-1066.

Egges, A., Kshirsagar, S., & Magnenat-Thalmann, N. (2003). A model for personality and emotion simulation. *Lecture Notes in Computer Science, 2773*, 453-461.

Epanechnikov, V. A. (1969). nonparametric estimation of a multivariate probability density. *Theory of Probability and Its Applications, 14*, 153-158.

Ericson, C. (2005). *Real-Time Collision Detection.* San Francisco, CA: Morgan Kaufmann Publishers.

Ester, M., Kriegel, H. P., Sander, J., & Xu, X. (1996). A density-based algorithm for discovering clusters in large spatial databases. *Proceedings of the 1996 International Conference on Knoledge Discovery and Data Mining*, (pp. 226-231). Portland, OR.

Ettema, D. (1996). Activity-based Travel Demand Modeling. *Ph.D. Thesis*, Technische Universiteit Eindhoven. Technische Universiteit Eindhoven, Netherlands.

Ettema, D., & Timmermans, H. J. (1997). Theories and models of activity-travel patterns. In D. Ettema, & T. H. P, *Activity-based Approaches to Travel Analyses* (pp. 1-36). Pergamon: Oxford.

Evans, S., Hudson-Smith, A., & Batty, M. (2005). 3-D GIS: Virtual London and beyond. *Cybergeo: European Journal of Geography*.

Fayyad, U., Piatetsky-Shapiro, G., & Smyth, P. (1996). From data mining to knowledge discovery - An review. In U. Fayyad, G. Piatetsky-Shapiro, P. Smyth, & U. R (Eds.), *Advances in knowledge discovery* (pp. 1-33). Cambridge, MA: AAAI Press/The MIT Press.

Feinberg, W. E., & Johnson, N. R. (2001). Primary group size and fatality risk in a fire disaster. *Proceedings of the Second International Symposium on Human Behavior in Fire, Understanding Human Behavior for Better Fire Safety Design* (pp. 11-22). London, UK: Interscience.

Festinger, L. (1957). *A theory of cognitive dissonance.* Stanford, CA: University Press.

Fisher, R. A. (1922). On the mathematical foundation of theoretical statistics. *Philosophical Transactions of the Royal Society of London, 222*, 309-368.

Flake, G. W. (2001). *The computational beauty of nature.* Cambridge, MA: The MIT Press.

Forsyth, D. (2009). *Group Dynamics* (5th ed.). Belmont, CA: Cenage Learning.

Fotheringham, A. S., & O'Kelly, M. E. (1989). *Spatial Interaction Models: Formulations and Applications.* Dordrecht: Kluwer Academic Publishers.

Franklin, S., & Graesser, A. (1996). Is it an agent, or just a program?: A taxonomy for autonomous agents. *Proceedings of the Workshop on Intelligent Agents III, Agent Theories, Architectures and Languages.* Springer-Verlag.

Fritz, H., Said, S., & Weimerskirch, H. (2003). Scale-dependent hierarchical adjustments of movement patterns in a long-range foraging seabird. *Proceedings of the Royal Society of London Series B*, *270*, pp. 1143-1148.

Fruin, J. J. (1993). The causes and prevention of crowd disasters. In R. A. Smith, & J. F. Dickie, *Engineering for Crowd Safety* (pp. 99-108). Amsterdam: Elsevier.

Fujita, M. (1982). Spatial patterns of residential development. *Journal of Urban Economics, 12*(1), 22-52.

Gaffney, S., & Smyth, P. (1999). Trajectory Clustering with Mixtures of Regression Models. *Proceedings of 5th ACM SIGKDD International Conference on on Knowledge Discovery and Data Mining*, (pp. 63-72). San Diego, CA.

Gaffney, S., Robertson, A., Smyth, P., Camargo, S., & Ghil, M. (2006). *Probabilistic Clustering of Extratropical Cyclones Using Regression Mixture Models.* Technical Report, UCI-ICS 06-02, University of California, Irvine.

Gahegan, M. (1999). Four barriers to the development of effective exploratory visualization tools for geosciences. *International Journal of Geographical Information Science, 13*, 289-309.

Gahegan, M., Wachowicz, M., Harrower, M., & Rhyne, T. M. (2001). The integration of geographic visualization with knowledge discovery in databases and geocomputation. *Cartography and Geographic Information Systems, 28*, 29-44.

Galea, E. R., & Galparsoro, J. M. (1993). *EXODUS: An evacuation model for mass transport vehicles.*

Gärling, T. (1994). Behavioral assumptions overlooked in travel-choice modelling. *Paper presented at the 7th international conference on travel behavior.* Valle, Chile.

Gärling, T. (1998). Behavioral assumptions overlooked in travel choice modeling. In J. D. Ortuzar, D. Hensher, & S. Jara-Diaz, *Travel Behavior Research: Updating the State of Play* (pp. 3-18). Amsterdam, Netherlands: Elsevier Science.

Gärling, T., & Gärling, E. (1988). Distance minimization in downtown pedestrian shopping. *Environment and Planning A, 20*(4), 547-554.

Gärling, T., Kalén, T., Romanus, J., Selart, M., & Vilhelmson, B. (1998). Computer simulation of household activity scheduling. *Environment and Planning A, 30*, 665-679.

Gatalsky, P., Andrienko, N., & G, A. (2004). Interactive analysis of event data using space-time cube. *Proceedings of the Eighth International Concerence on Information Visualization.*

Giannotti, F., & Pedreschi, D. (2007). *Mobility, Data Mining, and Privacy: Geographic Knowledge Discovery.* (F. Giannotti, & D. Pedreschi, Eds.) Berlin Heidelberg: Springer-Verlag.

Gipps, P. C. (1986). Simulation of pedestrian traffic in buildings. *Schriftenreihe des Instituts fuer Verkehrswesen, 35*. University of Karlsruhe.

Goerge, B., & Shekhar, S. (2006). Time aggregated graphs for modeling spatio-temporal networks. *Journal of Semantics of Data, 11*, 191-212.

Goldstein, M. L., Morris, S. A., & Yen, G. G. (2004). Problems with fitting to the power-law distribution. *European Physical Journal B, 41*, 255-258.

Golledge, R. G. (2004). Recent advances in human wayfinding and spatial cognition. *The Journal of the Institute of Electronics Information and Communication Enginneers, J87-A*(1), 3-12.

Golledge, R. G. (2008). Behavioral geography and the theoretical/quantitative revolution. *Geographical Analysis, 40*, 239-257.

Golledge, R. G., & Stimson, R. J. (1997). *Spatial behavior: A geographic perspective.* New York: The Guilford Press.

Golledge, R. G., & Timmermans, H. (1988). *Behavioral Modeling in Geography and Planning.* London, UK: Croom Helm.

Golledge, R. G., & Timmermans, H. (1990). Applications of behavioral research on spatial problems I: cognition. *Progress in Human Geography, 14*, 57-99.

Golledge, R. G., Rushton, G., & Clark, W. A. (1966). Some spatial characteristics of Iowa's dispersed farm population and their implications for the grouping of central place functions. *Geography, 42*, 261-272.

Golledge, R., Klatzky, R., & Loomis, J. (1996). Cognitive mapping and wayfinding by adults without vision. In *The Construction of Cognitive Maps* (pp. 215-246). Amsterdam: Springer.

González, M. C., Hidalgo, C. A., & Barabási, A. L. (2008). Understanding individual human mobility patterns. *Nature, 453*, 779-782.

Goodchild, M. F., & Mark, D. M. (1987). The fractal nature of geographic phenomena. *Annals of the Association of American Geographers, 77*(2), 265-278.

Grünwald, P., Myung, I. J., & Pitt, M. (2005). *Advances in Minimum Description Length: Theory and Applications.* MIT Press.

Guo, D., Liu, S., & Jin, H. (2010). A Graph-based Approach to Vehicle Trajectory Analysis. *Journal of Location Based Service, 4*(3), 183-199.

Guo, D., Peuquet, D., & Gahegan, M. (2003). ICEAGE: Interactive clustering and exploration of large and high-dimensional geodata. *Geoinformatica, 7*(3), 229–253.

Gustave, L. B. (1895). *The Crowd.* Dover Publications.

Guy, Y. (1986). Pedestrian Route Choice in Central Jerusalem. Beer Sheva, Hebrew: Department of Geography, Ben-Gurion University of The Negev.

Gwynne, S., Galea, E. R., Lawrence, P. J., Owen, M., & Filippidis, L. (1998). *Further validation of the BuildingEXODUS evacuation model using the Tsukuba dataset.* London, UK: University of Greenwich.

Gwynne, S., Galea, E. R., Owen, M., & Lawrence, P. J. (1998). *Validation of the BuildingEXODUS evacuation model.* London, UK: University of Greenwich.

Hägerstrand, T. (1970). What about people in regional science? *Papers of the Regional Science Association, 24*, 7-21.

Haken, H. (1983). *Synergetics, an Introduction* (3rd ed.). Berlin: Springer-Verlag.

Hall, M., E, F., Holmes, G., Pfahringer, B., Reutemann, P., & Witten, I. H. (2009). The WEKA Data Mining Software: An Update. *SIGKDD Explorations, 11*(1).

Hamacher, H. W., & Tjandra, S. A. (2001). Mathematical modelling of evacuation problems: a state of the art. In *Pedestrian and Evacuation Dynamics* (pp. 59-74). Berlin: Springer.

Hamnett, C. (1991). The blind men and the elephant: The explanation of gentrification. *Transactions of the Institute of British Geographers, 16*, 173-189.

Han, J., Lee, J.-G., & Kamber, M. (2009). An overview of clustering methods in geographic data analysis. In H. Miller, & J. Han (Eds.), *Geographic data mining and knowledge discovery* (2nd ed., pp. 149–185). CRC Press, Taylor and Francis Group.

Hansen, M. H., & Yu, B. (2001). Model selection and the principle of minimum description length. *Journal of the American Statistical Association, 96*(454), 746-774.

Harter, A. H., Steggles, P., Ward, A., & Webster, P. (1999). The anatomy of a context-aware application. *Proceedings of the 5th Annual ACM/IEEE International Conference on Mobile Computing and Networking* (pp. 59-68). Seattle, WA: ACM press.

Hartigan, J. A., & Wong, M. A. (1979). A K-means clustering algorithm. *Applied Statistics, 28*, 100-108.

Hazas, M., Scott, J., & Krumm, J. (2004). Location-aware computing comes of age. *Computer, 37*(2), 95-97.

Hedley, N. R., Drew, C. H., Arfin, E. A., & Lee, A. (1999). Hägerstrand revisited: Interactive space-time visualization of complex spatial data. *Informatica, 23*, 155-168.

Helbing, D. (1992). A Fluid-dynamic model for the movement of Pedestrians. *Complex systems, 6*, 391-415.

Helbing, D., & Johansson, A. (2007). Dynamics of crowd disasters: An empirical study. *Physical Review E*, 046109.

Helbing, D., & Johansson, A. (2010). Pedestrian, crowd and evacuation dynamics. *Encyclopedia of Complexity and Systems Science, 16*, 6476-6495.

Helbing, D., & Molnár, P. (1995). Social force model for pedestrian dynamics. *Physical Review E, 51*, 4282-4286.

Helbing, D., & Molnár, P. (1997). Self-organization phenomena in pedestrian crowds. In F. Schweitzer (Ed.), *Self organization of complex structures: From individual to collective dynamics.* CRC Press.

Helbing, D., Buzna, L., Johansson, A., & Werner, T. (2005). Self-organized pedestrian crowd dynamics: Experiments, simulations, and design solutions. *Transportation Science, 39*(1), 1-24.

Helbing, D., Farkas, I. J., Molnár, P., & Vicsek, T. (2002). Simulation of pedestrian crowds in normal and evacuation simulations. In M. Schreckenberg, & S. Sharma, *Pedestrian and Evacuation Dynamics* (pp. 21-58). Springer.

Helbing, D., Farkas, I., & Vicsek, T. (2000). Simulating dynamical features of escape panic. *Nature, 407*, 487- 490.

Helbing, D., Johansson, A., & Lämmer, S. (2007). Self-organization and optimization of pedestrian and vehicle traffic in urban environments. In S. Albeverio, D. Andrey, P. Giordano, & V. A, *The Dynamics of Complex Urban Systems - An Interdisciplinary Approach* (pp. 287-309). New York: Physica-Verlag.

Helbing, D., Keltsch, J., & Molnár, P. (1997). Modelling the evolution of human trail systems. *Nature, 388*, 47-50.

Helbing, D., Molnár, P., Farkas, I. J., & Bolay, K. (2001). Self-organizing pedestrian movement. *Environment and Planning B: Planning and Design, 28*, 361-383.

Henderson, A. (2007). *ParaView Guide: A Parallel Visualization Application.* Kitware Inc.

Henderson, L. F. (1971). The statistics of crowd fluids. *Nature, 229*, 381-383.

Hightower, J., & Borriello, G. (2001). A survey and taxonomy of location aware systems for ubiquitous computing. *IEEE Computer, 34*(8), 57-66.

Hightower, J., Vakili, C., Borriello, G., & Want, R. (2001). *Design and Calibration of the SpotON Ad-Hoc Location Sensing System.* Unpublished document.

Hightower, J., Want, R., & Borriello, G. (2000). *SpotON: An Indoor 3D Location Sensing Technology Based on RF Signal Strength.* UW CSE 2000-02-02, University of Washington, Seattle, WA.

Hill, M. R. (1982). Spatial structure and decision-making of pedestrian route selection through an urban environment. *Ph.D. Thesis*. University Microfilms International.

Hillier, B., & Hanson, J. (1984). *Social Logic of Space.* Cambridge University Press.

Hodges, J. S., & Dewar, J. A. (1992). *Is it You or Your Model Talking? A Framework for Model Validation.* Santa Monica, CA: Rand Corporation.

Holland, J. H. (1998). *Emergence: From chaos to order.* Reading, MA: Addison-Wesley Publishing Company, Inc.

Hoogendoorn, S. P. (2004). Walking behavior in bottlenecks and its implications for capacity. *Proceedings of the TRB 2004 Annual Meeting.*

Hoogendoorn, S. P., & Bovy, P. H. (2000). Gas-kinetic modeling and simulation of pedestian flows. *Transportation Research Record*(1710), 28-36.

Hoogendoorn, S. P., & Bovy, P. H. (2003). Simulation of pedestrian flows by optimal control and differential games. *Optimal Control Applications & Methods, 24*, 153-172.

Hoogendoorn, S. P., & Bovy, P. H. (2004). *Transportation Research Part B, 38*, 169-190.

Hoogendoorn, S. P., & Daamen, W. (2005). Pedestrian behavior at bottlenecks. *Transportation Science, 39*(2), 147-159.

Hoogendoorn, S. P., Daamen, W., & Landman, R. (2005). Microscopic calibration and validation of pedestrian models: Cross-comparison of models

using experimental data. In N. Waldau, P. Gattermann, H. Knoflacher, & M. Schreckenberg, *Pedestrian and evacuation dynamics 2005* (pp. 253-265). Berlin: Springer.

Hornsby, K., & Egenhofer, M. (2002). Modeling moving objects over multiple granularities. *Annals of Mathematics and Artificial Intelligence, 36*, 177-194.

Huang, Y., Shekhar, S., & Xiong, H. (2004). Discovering Co-location Patterns from Spatial Datasets: A General Approach. *IEEE Transactions on Knowledge and Data Engineering, 16*(12), 1472-1485.

Huang, Y., Xiong, H., Shekhar, S., & Pei, J. (2003). Mining Confident Co-location Rules without A Support Threshold. *Proceedings of the 18th ACM Symposium on Applied Computing*, (pp. 497-501).

Hughes, R. L. (2003). The flow of human crowds. *Annual Review of Fluid Mechanics, 35*, 169-182.

Isard, W., Azis, I. J., Drennen, M. P., Miller, R. E., Saltzmann, S., & Thorbecke, E. (1998). *Methods of Interregional and Regional Analysis.* Aldershot, UK: Ashgate.

Joh, C.-H., Arentze, T. A., & Timmermans, H. J. (2001). Understanding activity schduling and rescheduling behaviour: Theory and numerical illustration. *GeoJournal, 53*, 359-371.

Johansson, A. (2008). *Data-Driven Modeling of Pedestrian Crowds.* Technische Universität Dresden.

Johansson, A., & Helbing, D. (2008). From crowd dynamics to crowd safety: A video-based analysis. *Advances in Complex Systems, 11*(4), 497-527.

Johansson, A., Helbing, D., & Shukla, P. K. (2008). Specification of a microscopic pedestrian model by evolutionary adjustment to video tracking data. *Advances in Complex Systems, 10*(4), 271-288.

Johnson, P., Beck, D., & Horasan, P. (1994). Use of egress modeling in performance based fire engineering design - A fire safety study at the national gallery of Victoria. *Proceedings of the 4th International Symposium on Fire Safety Science*, (pp. 669-680). Ottawa, Canada.

Johnston, D. M., & Johnson, N. R. (1988). Role extension in disaster: Employee behavior at the Beverly Hills Supper Club Fire. *Sociological Focus, 22*(1), 39-51.

Jolliffe, I. T. (2002). *Principal component analysis* (2nd ed.). New York, NY: Springer-Verlag New York, Inc.

Kaiser, H. F. (1960). The application of electronic computers to factor analysis. *Educational and Psychological Measurement, 20*, 141-151.

Kanda, T., Glas, D. F., Shiomi, M., & Hagita, N. (2009). Abstracting People's Trajectories for Social Robots to Proactively Approach Customers. *IEEE Transactions on Robotics, 25*(6), 1382-1396.

Kapler, T., & Wright, W. (2004). GeoTime information visualization. *Proceedings of the IEEE Symposium on Information Visualization*, (pp. 25-32).

Kareiva, P. M., & Shigesada, N. (1983). Analysing insect movement as a correlated random walk. *Oecologia, 56*, 234-238.

Kaufman, A., & Mueller, K. (2005). Overview of volume rendering. In C. D. Hansen, & C. R. Johnson, *The Visualization Handbook* (pp. 127-174). Burlington, ON: Elsevier.

Kaufman, L., & Rousseeuw, P. J. (1990). *Finding groups in data: An introduction to cluster analysis.* John Wiley & Sons.

Kennedy, S. K., & Lin, W. (1986). FRACT-A fortran subroutine to calculate the variables necessary to determine the fractal dimension of closed forms. *Computers and Geoscience, 12*, 705-712.

Kentucky State Police. (1977). *Investigative Report to the Governor, Beverly Hills Supper Club Fire.* Frankfort, KY: Kentucky State Police.

Keßel, A., Klüpfel, H., Wahle, J., & Schreckenberg, M. (2001). Microscopic simulation of pedestrian crowd motion. In M. Schreckenberg, & S. D. Sharma, *Pedestrian and Evacuation Dynamics* (pp. 193-200). Berlin: Springer.

Kitamura, R. (1984). A model of daily time allocation to discretionary out-of-home activities and trips. *Transportation Research B, 18*, 255-266.

Kitamura, R. (1988). An evaluation of activity-based travel analysis. *Transportation, 15*, 9-34.

Kitamura, R., Nishii, K., & Goulias, K. (1990). Trip chaining behavior by central city commuters: a causal analysis of time-space constraints. In P. Jones, *Developments in Dynamic and Activity-based Approaches to Travel Analysis* (pp. 145-170). Brookfield, VT: Avebury.

Kitchin, R. M. (1994). Cognitive maps: waht are they and why study them? *Journal of Environmental Psychology, 14*(1), 1-19.

Knaden, M., & Wehner, R. (2003). Nest defense and conspecific enemy recognition in the desert ant Cataglyphis fortis. *Journal of Insect Behaviour, 16*(5), 717–730.

Knaden, M., & Wehner, R. (2004). Path integration in desert ants controls aggressiveness. *Science, 305*(5680), 60.

Knorr, E. M., & Ng, R. T. (1996). Finding aggregate proximity relationships and commonalities in spatial data mining. *IEEE Transactions on Knowledge and Data Engineering, 8*(6), 884–897.

Koozyt. (2008). Media Art and Technology Collaboration between Koozyt and Masayuki Akamatsu - "Location Amplifier for Tokyo Train" at Where2.0. Koozyt, Inc.

Koperski, K., & Han, J. (1995). Discovery of spatial association rules in geographic information databases. *Proceeding of the 4th International Symposium on Advances in Spatial Databases* (pp. 47-66). London, UK: Springer-Verlag.

Kraak, M.-J., & Koussoulakou, A. (2004). Visualization environment for the space-time cube. *Proceedings of the 11th International Symposium on Spatial Data Handling: Advances in Spatial Data Handling II*, (pp. 23-25).

Kritzler, M., Raubal, M., & Krüger, A. (2007). A GIS framework for spatio-temporal analysis and visualization of laboratory mice tracking data. *Transactions in GIS, 11*(5), 765-782.

Kuligowski, E. D. (2003). The evaluation of a performance-based design process for a hote building: The comparison of two egress models. University of Maryland.

Kwan, M.-P. (1998a). Interactive geovisualization of activity-travel patterns using 3D geographical information systems: A methodological exploration with a large data set. *Transportation Research Part C, 8*, 185-203.

Kwan, M.-P. (1998b). Space-time and integral measures of individual accessibility: A comparative analysis using a point-based framework. *Geographical Analysis, 30*(3), 191-216.

Kwan, M.-P. (1999). Gender and individual access to urban opportunities: A study using space-time measures. *Professional Geographer, 51*(2), 210-227.

Kwan, M.-P. (2000a). Interactive geovisualization of activity-travel patterns using three-dimensional geographical information systems: a methodological exploration with a large data set. *Transportation Research Part C, 8*, 185-203.

Kwan, M.-P. (2000b). Analysis of human spatial behavior in a GIS environment: Recent developments and future prospects. *Journal of Geographical Systems, 2*, 85-90.

Kwan, M.-P., & Hong, X. (1998). Network-based constraints-oriented choice set formation using GIS. *Geographical Systems, 5*, 139-162.

Latane, B., & Darley, J. M. (1970). *The Unresponsive Bystander: Why doesn't he help?* New York: Appleton-Centry Crofts.

Laube, P., & Purves, R. S. (2006). An Approach to Evaluating Motion Pattern Detection Techniques in Spatio-temporal Data. *Computers, Environment and Urban Systems, 30*, 347-374.

Laube, P., Dennis, T., Forer, P., & Walker, M. (2007). Movement beyond the snapshot: Dynamic analysis of geospatial lifelines. *Computers, Environment and Urban Systems, 31*, 481-501.

Laube, P., Imfeld, S., & Weibel, R. (2005). Discovering relative motion patterns in groups of moving point objects. *Journal of Geographic Information Science, 19*, 639-668.

LeBon, G. (1895). *The Crowd: A Study of the Popular Mind.* London, UK: Transaction Publishers.

Ledoux, H., & Gold, C. M. (2006). A voronoi-based map algebra. In A. Reidl, W. Kainz, & G. Elmes, *Progress in Spatial Data Handling - 12th International Symposium on Spatial Data Handling* (pp. 117-131). Springer.

Lee, J. G., Han, J., & Li, X. (2008). Trajectory outlier detection: a partition-and-detect framework. *IEEE 24th International Conference on Data Engineering (ICDE)*, (pp. 140-149). Cancun.

Lee, J. G., Han, J., & Whang, K. Y. (2007). Trajectory clustering: A partition-and-group framework. *Proceedings of the ACM SIGMOD International Conference on Management of Data*, (pp. 593-604). Beijing, China.

Lenntorp, L. (1976). *Paths in Space-Time Environment: a Time-Geographic Study of the Movement Possibilities of Individuals.* The Royal University of Lund, Department of Geography, Lund, Sweden.

Levoy, M. (1988). Volume rendering: Display of surfaces from volume data. *IEEE Computer Graphics and Applications, 8*, 29-37.

Lloyd, S. P. (1982). Least Squares Quantization in PCM. *IEEE Trans. Information Theory, 28*, 128-137.

Locher, D. A. (2002). *Collective Behavior.* Upper Saddle River, NJ: Prentice Hall.

Longley, P. A., Goodchild, M. F., Maguire, D. J., & Rhind, D. W. (2001). *Geographic Information Systems and Science.* New York, NY: John Wiley & Sons, Ltd.

Louviere, J. J., Hensher, D. A., & Swait, J. D. (2000). *Stated choice methods.* Cambridge, UK: Cambridge University Press.

MacQueen, J. (1967). Some methods for classification and analysis of multivariate observations. *Proceedings of the Berkeley Symposium on Mathematical Statistics and Probability*, (pp. 281-297).

Mandelbrot, B. (1967). How long is the coast of Britain? Statistical self-similarity and fractional dimension. *Science, 156*, 636-638.

Mandelbrot, B. (1983). *The Fractal Geometry of Nature.* San Francisco, CA: Freeman.

Manson, S. M. (2007). Challenges in evaluating models of geographic complexity. *Environment and Planning B: Planning and Design, 34*, 245-260.

Mardia, K. V., & Jupp, P. E. (2000). *Directional Statisitics.* Wiley series.

Mark, E., & Wright, R. (2005). Assimilation and differences between the settlement patterns of individual immigrants and immigrant households. *Proceedings of the National Academy of Sciences, 102*(43), 15325-15330.

McCarthy, K. (1982). An analytical model of housing search and mobility. In W. A. Clark (Ed.), *Modelling Housing Market Search* (pp. 30-53). London, UK: Croom Helm Ltd.

McFadden, D. L. (1974). Conditional logit analysis of qualitative choice behavior. In P. Zarembka, *Frontiers in Econometrics* (pp. 105-142). New York: Academic Press.

Mchugh, K. E., & Gober, P. (1992). Short-term dynamics of the U.S. interstate migration system, 1980-1988. *Growth and Change, 23*(4), 428-445.

Mclntosh, J., & Yuan, M. (2005). Assessing similarity of geographic processes and events. *Transactions in GIS, 9*(2), 223-245.

McPhail, C. (1994). The dark side of purpose: Individual and collective violence in riots. *The Sociological Quarterly, 35*(1), 1-32.

McPhail, C., & Tucker, C. W. (1992). Simulating individual and collective action in temporary gatherings. *Social Science Computer Review, 10*, 1-28.

Mennis, J., & Guo, D. (2009). Spatial data mining and geographic knowledge discovery - An introduction. *Computers, Environment, and Urban Systems, 33*, 403-408.

Mennis, J., Viger, R., & Tomlin, C. D. (2005). Cubin map algebra functions for spatio-temporal analysis. *Cartography and Geographic Information Science, 32*(1), 17-32.

Mileti, D. S., & Sorensen, J. H. (1990). *Communication of emergency public warnings, a social science* .

Miller, H. J. (1991). Modeling accessibility using space-time prism cocepts within geographic information systems. *International Journal of Geographic Information Systems, 5*, 287-301.

Miller, H. J. (1999). Measuring space-time accessibility benefits within transportation networks: Basic theory and computational procedures. *Geographical Analysis, 31*, 187-212.

Miller, H. J. (2003). What about people in geograhic information science? *Computers, Environment, and Urban Systems, 27*, 227-453.

Miller, H. J. (2004). Activities in space and time. In P. Stopher, K. Button, K. Haynes, & D. Hensher, *Handbook of Transport: Transport Geography and Spatial Systems* (Vol. 5). Elsevier Science.

Miller, H. J. (2005). A measurement theory for time geography. *Geographical Analysis, 37*, 17-45.

Miller, H. J., & Bridwell, S. A. (2009). A field-based theory for time geography. *Annals of the Association of American Geographers, 99*(1), 49-75.

Miller, H. J., & Han, J. (2001). Geographic data mining and knowledge discovery: An overview. In H. J. Miller, & J. Han (Eds.), *Geographic data mining and knowledge discovery* (pp. 3-32). Taylor & Francis.

Miller, H. J., & Han, J. (2009). Geographic data mining and knowledge discovery: An overview. In H. J. Miller, & H. J (Eds.), *Geographic data mining and knowledge discovery* (2nd ed., pp. 1–26). Taylor & Francis.

Mishra, A. R. (2004). *Fundamentals of cellular network planning and optimisation.* John Wiley and Sons.

Montello, D. R. (2001). Spatial cognition. In N. J. Smelser, & P. B. Baltes, *International Encyclopedia of the Social & Behavioral Sciences* (pp. 14771-14775). Oxford: Pergamon Press.

Montello, D. R. (2005). Navigation. In P. Shah, A. Miyake, P. Shah, & A. Miyake (Eds.), *The Cambridge handbook of visuospatial thinking* (pp. 257-294). New York, NY: Cambridge University Press.

Montello, D. R. (2009). Cognitive geography. In R. Kitchin, N. Thrift, R. Kitchin, & N. Thrift (Eds.), *International encyclopedia of human geography* (Vol. 2, pp. 160-166). Oxford Elsevier Science.

Moore, A., Whigham, P. A., Holt, A., Aldridge, C. H., & Hodge, K. (2003). Sport and time geography: a good match? *15th Annual Colloquium of the Spatial Information Research Centre*, (pp. 109-116). Dunedin, New Zealand.

Moussaid, M., Perozo, N., Garnier, S., Helbing, D., & Theraulaz, G. (2010). The walking behavior of pedestrian social groups and its impact on crowd dynamics. *Plos One*, e10047.

Muhdi, R. A. (2006). Evacuation modeling: Development, characteristic, and limitations. *Proceedings of the IEEE CEC*, (pp. 87-92). Vancouver, BC.

Muramatsu, M., Irie, T., & Nagatani, T. (1999). Jamming transition in pedestrian counter flow. *Physica A: Statistical Mechanics And Its Application, 267*, 487-4998.

Nakaya, T., & Yano, K. (2008). Spatio-temporal three-dimensional mapping of crime events: Visualising spatio-temporal clusters of snatch-and-run offences. *Journal of Geography, 117*(2), 506-521.

Nakaya, T., & Yano, K. (2010). Visualising crime clusters in a space-time cube: An exploratory data-analysis approach using space-time kernel density estimation and scan statistics. *Transactions in GIS, 14*(3), 223-239.

Nams, V. O. (2005). Using animal movement paths to measure response to spatial scale. *Oecologia, 143*, 179-188.

Nams, V. O., & Bourgeois, M. (2004). Fractal dimension measures habitat use at different spatial scales: an example with marten. *Canadian Journal of Zoology, 82*, 1738-1747.

Nara, A., & Torrens, P. M. (2007). Spatial and temporal analysis of pedestrian egress behavior and efficiency. In H. Samet, C. Shahabi, & M. Schneider, *Association of Computing Machinery (ACM) Advances in Geographic Information Systems* (pp. 284-287). New York: ACM.

Nara, A., Izumi, K., Iseki, H., Suzuki, T., Nambu, K., & Sakurai, Y. (2009). Trajectory data mining for surgical workflow analysis. *Proceedings of GeoComputation 2009.*

Newman, M. E. (2005). Power laws, Pareto distributions and Zipf's law. *Contemporary Physics, 46*, 323-251.

Ng, R. T., & Han, J. (1994). Efficient and effective clustering methods for spatial data mining. *Proceedings of the 20th Very Large Database Conference*, (pp. 144-155).

Olguín, D. O., & Pentland, A. (2006). Human activity recognition: Accuracy across common locations for wearable sensors. *IEEE 10th International Symposium on Wearable Computers.* Montreaux, Switzerland.

Openshaw, S. (1983). The modifiable areal unit problem. *CATMOG 83.*

Orr, R. J., & Abowd, G. D. (2000). The Smart Floor: Amechanism for natural user identificaiton and tracking. *Proceedings of the 2000 Conference on Human Factors in Computing Systems.*

O'Sullivan, D. (2002). Toward microscale spatial modeling of gentrification. *Journal of Geographical Systems, 4*, 251-274.

O'Sullivan, D. (2004). Complexity science and human geography. *Transactions of the Institute of British Geographers, 29*(3), 282-295.

O'Sullivan, D., Manson, S. M., Messina, J. P., & Crawford, T. W. (2006). Space, place, and complexity science. *Environment and Planning A, 38*, 611-617.

Pan, X., Han, C. S., Dauber, K., & Law, K. H. (2006). Human and Social Behavior in Computational Modeling and Analysis of Egress. *Journal of automation in construction, 15*, 448-461.

Pari, F., Luschi, P., Akesson, S., Capogrossi, S., & Hays, G. C. (2000). Open-sea migraition of magnetically disturbed sea turtles. *The journal of experimental biology, 203*, 3435-3443.

Pas, E. (1988). Weekly travel-activity behavior. *Transportation Research, 13*, 89-109.

Pas, E. (1990). Is travel demand analysis in the doldrums? In P. Jones, *Developments in Dynamic and Activity-based Approaches to Travel Analysis* (pp. 3-27). Brookfield, VT: Avebury.

Pas, E., & Koppelman, F. (1987). An examination of the determinants of day-to-day variability in individuals' urban travel behavior. *Transportation, 13*, 183-200.

Patterson, D., Liao, L., Fox, D., & Kautz, H. (2003). Inferring high-level behavior from low-level sensors. *Ubicomp*, 73-89.

Pelechano, N., Allbeck, J. M., & Badler, N. I. (2007). Controlling individual agents in high-density crowd simulation. *Proceedings of ACM SIGGRAPH / Eurographics Symposium on Computer Animation* (pp. 99-108). San Diego, CA: ACM Press.

Pelechano, N., Allbeck, J., & Badler, N. I. (2008). *Virtual Crowd: Methods, Simulation, and Control.* Morgan & Claypool Publishers.

Pelechano, N., O'Brien, K., Silverman, B., & Badler, N. (2005). Crowd simulation incorporating agent psychological models, roles and communication. *Proceedings of 1st International Workshop on Crowd Simulation*, (pp. 21-30).

Penn, A., & Turner, A. (2002). Space syntax based agent simulation. *Proceedings of the 1st International Conference on Pedestrian and Evacuation Dynamics* (pp. 99-114). University of Duisburg, Germany: Springer: Berlin.

Peuquet, D. J. (2002). *Representation of space and time.* New York, NY: The Guilford Press.

Portugali, J. (2000). *Self-organization and the city.* Berlin: Springer-Verlag.

Portugali, J. (2006). Complexity theory as a link between space and place. *Environment and Planning A, 38*, 647-664.

Portugali, J., Benenson, I., & Omer, I. (1994). Socio-spatial residential dynamics: Stability and instability within a self-organizing city. *Geographical Analysis, 26*, 321-340.

Portugali, J., Benenson, I., & Omer, I. (1997). Spatial cognitive dissonance and sociospatial emergence in a self-organizing city. *Environment and Planning B: Planning and Design, 24*, 263-285.

Powers, W. T. (1973). *Behevior: The control of perception.* Chicago, IL: Aldine.

Prigogine, I. (1980). *From Being to Becoming.* San Francisco, CA: Freeman & Co.

Priyantha, N. B., Chakraborty, A., & Balakrishnan, H. (2000). The cricket location-support system. *Proceedings of MOBICOM* (pp. 32-43). Boston, MA: ACM press.

Proulx, G. (2002). Movement of people: The evacuation timing. In P. J. Dinenno, & W. D. Walton, *The SFPE Handbook of Fire Protection Engineering* (Third ed., pp. 3-341-3-366). Bethesda, MD: Society of Fire Protection Engineers.

Proulx, G., Latour, J. C., McLaurin, J. W., Pineau, J., Hoffman, L. E., & Laroche, C. (1995). *Housing evacuation of mixed abilities occupants in highrise buildings.* Internal Report, No.706, The Institute for Research in Construction, Canada.

Pullar, D. (2001). MapScript: A map algebra programming language incorporating neighborhood analysis. *GeoInformatica, 5*(2), 145-163.

Quinlan, J. R. (1986). Induction of decision trees. *Machine Learning, 1*, 81-106.

Quinlan, J. R. (1993). *C4.5: Program for Machine Learning.* Morgan Kaufmann Publishers.

Quinlan, J. R. (1996). Learning Decision Tree Classifiers. *ACM Computing Surveys, 28*(1), 71-72.

R Development Core Team. (2008). *R: A language and environment for statistical computing.* Vienna, Austria: R Foundation for Statistical Computing.

Ramos-Fernandez, G., Mateos, J. L., Miramontes, O., Cocho, G., Larralde, H., & Ayala-Orozco, B. (2003). Levy walk patterns in the foraging movements of spider monkeys. *Behavioral Ecology and Sociobiology, 55*(2), 223-230.

Reades, J., Calabrese, F., & Ratti, C. (2009). Eigenplaces: analysing cities using the space-time structure of teh mobile phone network. *Environment and Planning B: Planning and Design, 36*, 824-836.

Recker, W. W., McNally, M. G., & Roth, G. S. (1986). A model of complex travel behavior: an operational model. *Transportation Research A, 20*, 319-330.

Rekimoto, J., Shionozaki, A., Sueyoshi, T., & Miyaki, T. (2006). PlaceEngine: a WiFi location platform based on realworld folksonomy. *Internet Conference 2006*, (pp. 95-104). Tokyo, JP.

Reynolds, C. W. (1987). Flocks, Herds, and Schools: A Distributed Behavioral Model. *Proceedings of SIGGRAPH '87 Conference*, *25(4)*, pp. 25-34.

Reynolds, C. W. (1999). Behaviors For Autonomous Characters. *Proceedings of Game Developers Conference* (pp. 763-782). San Francisco, CA: Miller Freeman Game Group.

Sadalla, E. K., & Montello, D. R. (1989). Remembering changes in direction. *Environment and Behavior, 21*(3), 346-363.

Sakuma, T., Mukai, T., & Kuriyama, S. (2005). Psychological model for animating crowded pedestrians. *Journal of Visualization and Computer Animation, 16*(3-4), 343-351.

Sawyer, R. K. (2002). Nonreductive individualism, part I: supervenience and wild disjunction. *Philosophy of the Social Sciences, 32*, 537-559.

Schadschneider, A., & Seyfried, A. (2009). Empirical results for pedestrian dynamics and their implications for celluar automata models. In H. Timmermans, *Pedestrian Behavior: Data Collection and Applications* (pp. 27-43). Emerald Group Publishing Limited.

Schadschneider, A., Klingsch, W., Klüpfel, H., Kretz, T., Rogsch, C., & Seyfried, A. (2008). Evacuation dynamics: Empirical results, modeling and analysis. In B. Meyers, *Encyclopedia of Complexity and System Science.* Berlin: Springer.

Schelhorn, T., O'Sullivan, D., Haklay, M., & Thurstain-Goodwin, M. (1999). STREETS: An Agent-Based Pedestrian Model. *Centre for Advanced Spatial Analysis, Working Paper Series, Paper 9.*

Schelling, T. C. (1971). Dyamic models of segregation. *Journal of Mathematical Sociology, 1*(1), 143-186.

Schelling, T. C. (1974). On the ecology of micro-motives. In R. Marris (Ed.), *The corporate Society.* London, UK: Macmillan.

Schmitt, F. G., & Seuront, L. (2001). Multifractal random walk in copepod behaviour. *Physica A, 301*(1-4), 375–396.

Schultz, D. P. (1964). *Panic Behavior, Discussion and Reading.* New York: Random House.

Schweingruber, D. (1995). A computer simulation of a sociological experiment. *Social Science Computer Review, 13*, 351-359.

Schweingruber, D., & Wohlstein, R. T. (2005). The madding crowd goes to school: Myths about crowds in introductory sociology textbooks. *Teaching Sociology, 33*(2), 136-153.

Scott, D. W. (1992). *Multivariate Density Estimation: Theory, Practice, and Visualization.* New York: John Wiley and Sons.

Sears, F. W., Zemansky, M. W., & Young, H. D. (1987). *University Physics.* (7th ed.). Reading, MA: Addison-Wesley.

Senevarante, P. N., & Morall, J. F. (1986). Analysis of factors affecting the choice of route of pedestrians. *Transportation Planning and Technology, 10*, 147-159.

Seyfried, A., Boltes, M., Kähler, J., Klingsch, W., Portz, A., Rupprecht, T., et al. (2010). Enhanced empirical data for the fundamental diagram and the flow through bottlenecks. In W. W. Klingsch, C. Rogsch, A. Schadschneider, & M. Schreckenberg, *Pedestrian and Evacuation Dynamics 2008* (pp. 145-156). Springer.

Seyfried, A., Rupprecht, T., Passon, O., Steffen, B., Klingsch, W., & Boltes, M. (2007). New insights into pedestrian flow through bottlenecks. *Physics*.

Seyfried, A., Steffen, B., Klingsch, W., & Boltes, M. (2005). The fundamental diagram of pedestrian movement revisited. *Journal of Statistical Mechanics*, P10002.

SFPE. (2003). *Engineering Guide: Human Behavior in Fire.* Bethesda, MD: Society of Fire Protection Engineers.

Shaw, S.-L., & Yu, H. (2009). A GIS-based time-geographic approach of studying individual activities and interactions in a hybrid physical-virtual space. *Journal of Transport Geography, 17*(2), 141-149.

Shaw, S.-L., Yu, H., & Bombom, L. S. (2008). A space-time GIS approach to exploring large individual-based spatiotemporal datasets. *Transactions in GIS, 12*(4), 425-441.

Shen, Z., & Ma, K.-L. (2008). MobiVis: A visualization system for exploring mobile data. *IEEE PacificVIS*, (pp. 175-182).

Shen, Z., Ma, K.-L., & Eliassi-Rad, T. (2006). Visual analysis of large heterogeneous social networks by semantic and structural abstraction. *IEEE Transactions on Visualization and Computer Graphics, 12*(6), 1427-1439.

Silverman, B. W. (1986). *Density Estimation for Statistics and Data Analysis.* London: Chapman & Hall.

Sime, J. D. (1989). Handicapped people or handicapping environments? *Building Journal Hong Kong*, 84-92.

Sime, J. D., & Kimura, M. (1988). The timing of escape: Exit choice behaviour in fires and building evacuation. In J. Sime, & F. N. Spon, *Safety in the Built Environment* (pp. 48-61). London.

Sims, D. W., Southall, E. J., Humphries, N. E., Hays, G. C., Bradshaw, C. J., Pitchford, J. W., et al. (2008). Scaling laws of marine predator search behavior. *Nature, 451*, 1098-1103.

Skyhook. (2008). *Locr and Skyhook Wireless to Jumpstart Geotagging*. Retrieved 4 10, 2011, from http://www.skyhookwireless.com/press/skyhook_locr.php

Smith, D. R., Brown, B. B., Yamada, I., Kowaleski-Jones, L., Zick, C., & Fan, J. X. (2008). Walkability and Body Mass Index: Density, Design, and New Diversity Measures. *American Journal of Preventive Medicine, 35*(3), 237-244.

Smith, M. J., Goodchild, M. F., & Longley, P. A. (2009). *Geospatial Analysis - A Comprehensive Guide to Principles, Techniques, and Software Tools* (3rd ed.). SPRINT.

Smith, N. (1979). Toward a theory of gentrification: A back to the city movement by capital not people. *Journal of the American Planning Association, 45*, 538-548.

Still, G. K. (2000). *Crowd dynamics.* Ph.D. thesis, University of Warwick, UK.

Takahashi, K., Tanaka, T., & Kose, S. (1988). An evacuation model for use in fire safety designing of building. *Proceedings of the 2nd International Symposium on Fire Safety Science*, (pp. 551-560).

Takeyama, M., & Couclelis, H. (1997). Map Dynamics: Integrating Cellular Automata and GIS through Geo-Algebra. *International Journal of Geographic Information Systems, 11*(1), 73-91.

Thalmann, D., & Musse, S. R. (2007). *Crowd Simulation.* London, UK: Springer-Verlag.

Thrift, N. (1999). The Place of Complexity. *Theory, Culture & Society, 16*(3), 31-69.

Thünen, J. v. (1826). *Der Isolierte Staat in Beziehung auf Landtschaft und Nationalökonomie.* (C. M. Wartenberg, Trans.) Pergamon Press.

Tibshirani, R., Walther, G., & Hastie, T. (2001). Estimating the number of clusters in a dataset via the Gap Statistics. *Journal of the Royal Statistic Society: B, 2*, 411-423.

Timmermans, H. (2009). *Pedestrian Behavior: Models, Data Collection and Applications.* Bingley, UK: Emerald Group Publishing Ltd.

Timmermans, H., & Arentze, C. H. (2002). Analysing space-time ehaviour: new approaches to old problems. *Progress in Human Geography, 26*(2), 175-190.

Tomlin, C. D. (1990). *Geographic Information Systems and Cartographic Modeling.* Englewood Cliffs, NJ: Prentice-Hall.

Tomlin, C. D., & Berry, J. K. (1979). A mathematical structure for cartographic modeling in environmental analysis. *Proceedings of the American Congress on Surveying and Mapping*, (pp. 269-283).

Torrens, P. M. (2000). *How land-use transportation models work.* Working Paper, 28, University College London, Centre for Advanced Spatial Analysis, London, UK.

Torrens, P. M. (2002). Cellular automata and multi-agent systems as planning support tools. In S. Geertman, & J. Stillwell (Eds.), *Planning Support Systems in Practice* (pp. 205-222). London, UK: Springer-Verlag.

Torrens, P. M. (2003). *Simulating sprawl: A dynamic entity-based approach to modeling North American suburban sprawl using cellular automata and multi-agent systems.* Ph.D. thesis, University College London, London, UK.

Torrens, P. M. (2011). Moving agent pedestrians through space and time. *Annals of the Association of American Geographers*, (in press).

Torrens, P. M., & Benenson, I. (2005). Geographic Automata Systems. *International Journal of Geographic Information Science, 19*(4), 385-412.

Torrens, P. M., & Li, X. (2011). Building agent-based walking models by machine-learning on diverse databases of space-time trajectory samples. *Transactions in Geographic Information Science*, in press.

Treuille, A., Cooper, S., & Popovic, Z. (2006). Continuum crowds. *ACM Transactions on Graphics, 25*(3), 1160-1168.

Treuille, A., Cooper, S., & Popović, Z. (2006). Continuum crowds. *ACM Transactions on Graphics, 25*(3), 1160-1168.

Tucker, C. W., Schweingruber, D., & McPhail, C. (1999). Simulating arcs and rings in gatherings. *International Journal of Human-Computer Studies, 50*, 581-588.

Tucker, C. W., Schweingruber, D., & McPhail, C. (1999). Simulating arcs and rings in gatherings. *International Journal of Human-Computer Studies, 50*, 581-588.

Turchin, P. (1998). *Quantitative Analysis of Movement: Measuring and Modeling Population Redistribution in Animals and Plants.* Sunderland, MA: Sinauer Associates.

Turner, R. H., & Killian, L. M. (1972). *Collective Behavior* (2nd ed.). Prentice-Hall.

Turner, R. H., & Killian, L. M. (1987). *Collective Behavior.* Englewood Cliffs, NJ: Prentice-Hall.

Walmsley, D. J., & Lewis, G. J. (1984). *Human Geography: Behavioral Approaches.* London, UK: Longman.

Walton, D., & Sunseri, S. (2010). Factors influencing the decision to drive or walk short distances to public transport facilities. *International Journal of Sustainable Transportation, 4*(4), 212-226.

Wang, W., Yang, J., & Muntz, R. (1997). STING: A Statistical Information Grid Approach to Spatial Data Mining. *Proceedings of the 23rd Very Large Databases Conference*, (pp. 186-195). Athens, Greece.

Want, R., & Russell, D. M. (2000). Ubiquitous electronic tagging. *IEEE Distributed Systems Online, 1*(2).

Want, R., Hopper, A., Falcao, V., & Gibbons, J. (1992). The active badge location system. *ACM Transactions on Information Systems, 10*(1), 91-102.

Ware, C., Arsenault, R., Plumlee, M., & Wiley, D. (2006). Visualizing the underwater behavior of humpback whales. *IEEE Computer Graphics and Applications*, 14-18.

Weibel, R., Sack, J. R., Sester, M., & Bitterlich, W. (2008). *Representation, Analysis and Visualization of Moving Objects.* Short report of Dagstuhl Seminar.

Weimerskirch, H., Bonadonna, F., Bailleul, F., Mabille, G., Dell'Omo, G., & Lipp, H. (2002). GPS Tracking of foraging albatrosses. *Science, 295*, 1259.

Wheler, L. (1966). Toward a theory of behavioral contagion. *Psychological Review, 73*, 179-192.

Willems, N., Wetering, H. v., & Wijk, J. J. (2009). Visualization of vessel movements. *IEEE VGTC Symposium on Visualization, 28*(3), 959-966.

Williamson, D., McLafferty, S., Goldsmith, V., Mollenkopf, J., & P, M. (1999). *A better method to smooth crime incident data*. Retrieved November 1, 2010, from ESRI ArcUser Magagine: http://www.esri.com/news.arcuser/0199/crimedata.html

Wilson, A. G. (1975). *Urban and regional models in geography and planning.* London, UK: John Wiley & Sons.

Wilson, A. G. (1981). *Catastrophe Theory and Bifurcation.* London, UK: Croom Helm.

With, K. A. (1994). Ontogenetic shift in how grasshoppers interact with landscape structure: An analysis of movement patterns. *Functional Ecology, 8*, 477-485.

Witten, T. A., & Sander, L. M. (1981). Diffusion-Limited Aggregation, a Kinetic Critical Phenomenon. *Physics Review Letters, 47*, 1400-1403.

Wolfson, O. (2002). Moving objects information management: The database challenge. *The proceedings of the 5th workshop on Next Generation Information Technology and Systems (NGITS'2002)*, (pp. 75-89). Caesarea, Israel.

Wrigley, N., Holt, T., Steel, D., & Tranmer, M. (1996). Analysing, modeling, and resolving the ecological fallacy. In P. A. Longley, & M. Batty (Eds.), *Spatial analysis: Modelling in a GIS environment* (pp. 25-40). Cambridge: GeoInformation International.

Yoshida, T., Shoda, R., & Motoda, H. (2006). Graph clustering based on structural similarity of fragments. In K. P. Jantke, A. Lunzer, N. Spyratos, & Y. Tanaka (Eds.), *Federation over the web* (Vol. 3847, pp. 97-114).

Yu, H. (2006). Spatio-temporal GIS design for exploring interactions of human activities. *Cartography and Geographic Information Science, 33*, 3-19.

Yu, H., & Shaw, S.-H. (2007). Revisiting Hagerstrand's time-geographic framework for individual activities in the age of instant access. In H. J. Miller (Ed.), *Societies and Cities in the Age of Instant Acess* (pp. 103-118). Dordrecht, Netherlands.: Springer.

Yu, H., & Shaw, S.-L. (2008). Exploring potential human activities in physical and virtual spaces: A spatio-temporal GIS approach. *International Journal of Geographic Information Science, 22*, 409-430.

Yuan, M. (2001). Representing complex geographic phenomena in GIS. *Cartography and Geographic Information Science, 28*(2), 82-96.

Yuan, M., Mark, D., Egenhofer, M., & Peuquet, D. (2004). Extensions to geographic representation. In R. McMaster, & E. Usery, *A Research Agenda for Geographic Information Science* (pp. 129-156). Boca Raton, FL: CRC Press.

Zhang, T., Ramakrishnan, R., & Livny, M. (1996). BIRCH: An efficient data clustering method for very large database. *Proceedings of the ACM SIGMOD International Conference on Management of Data*, (pp. 103-114).