

# Video Trajectory Analysis

Jiang Bian

Faculty of Engineering and Information Technology

University of Technology Sydney

A thesis submitted for the degree of

*Doctor of Philosophy*

2019



To my loving parents  
*Zuosen Bian* and *Yanhui He*  
my wife  
*Yiran Ding*  
and my children  
*Xuyuan Bian* and *Jingbo Bian*



## **Certificate of Original Authorship**

I, Jiang Bian declare that this thesis, is submitted in fulfilment of the requirements for award of Doctor of Philosophy, in the Faculty of Engineering and Information Technology at the University of Technology Sydney.

This thesis is wholly my own work unless otherwise reference or acknowledged. In addition, I certify that all information sources and literature used are indicated in the thesis.

This document has not been submitted for qualifications at any other academic institution.

This research is supported by an Australian Government Research Training Program.

Jiang Bian



## Acknowledgements

I would like to express my special appreciation and thanks to my supervisor Professor Dacheng Tao and Professor Maolin Huang. Dacheng accepted my application four years ago and gave me the chance to change my life. In first 45 months, he consistently taught me how to research. From reading literatures to writing my own papers, from implementing others' methods to devising my own methods, I learned a lot under his instructions. Maolin gave me the chance to finish my PhD degree and taught me how to write papers and submit one conference paper to review under his supervision. He gave his total support to help me improve myself.

I have to say thank you to my wife Yiran Ding, she tried her best to support me and gave me all her have to help me finish my degree without disturbing. I also wish to give special thanks to Chaoyue Wang, Guoliang Kang, Dayong Tian and Jiayan Qiu. Without their helps, my life in Sydney would not be so easy. I am grateful to Zijing Chen. Our work stations are neighboring. Her opinions deeply impressed me and gave me more space to explore. I would like to give my gratitude to Jun Li and Maoying Qiao for their helps on my research.

I would like to give thanks to my friends I met in Sydney, especially to Zhenkai Hao. Zhenkai is an graduated student from University of Sydney. He shared lots of his experiences on learning English to me, and life skills.

Finally, I would like to express my gratitude to my family, my parents and my children, for their encouragement and support.

## Abstract

Considering the critical role of trajectory data mining in modern intelligent systems for surveillance security, abnormal behavior detection, crowd behavior analysis, and traffic control. Furthermore, with the widely spreading of camera, trajectories are recorded by camera, so trajectory analysis including trajectory clustering in computer vision is of great use for a lot of works. However, video trajectories analysis is also a hard work, because its limited information to generate trajectories and few representation methods are available. Thus, the better performance could be reached if more reliable motion information is employed. A lot of characterizations are contained in trajectory data that can be useful and powerful in trajectory clustering including distance, speed, direction, relative displacement and some other features. Finally, in the case that a large number of trajectory data need to be cluster into small number of categorizes which are hidden “groups”, an unsupervised clustering model is also required to implement the goal.

In addition, with more and more lecture videos are available on the Internet, on-line learning and e-learning are getting increasing concerns because of many advantages such as high degree of interactivity. The semantic content discovery for lecture video is very important. However, every lecture video contains a lot of semantic information including spoken language and lecture notes, so how to use all these features is a key problem to improve the performance of e-learning.

Therefore, a novel method is proposed in this paper. Reference points are detected and the scale-invariant feature transform (SIFT) descriptor is used to represent the image patches around the points. In addition, SIFT is a descriptor that is fast and robust to match. In order



to unify the lengths of trajectories, Discrete Fourier Transformation (DFT) transforms trajectories into frequency domain with a fixed length, so that pattern information is retained. Furthermore, one more feature type is involved to describe object motion that presents the motion of object relative to the camera, and the difference between the static objects and moving objects can be figured out.

Latent Dirichlet Allocation (LDA) has great performance on natural language processing, but it prefers to model discrete words only. However, another different kind of semantic feature, continues feature, involves in, so we derive a novel clustering model called derived LDA model which the word-topic distribution following Multivariate distribution. After derived LDA, we derive dual-variable LDA model that processes two different features parallel. Furthermore, a detailed derivative process is given to support our model.

In the experiment, we applied our model into two data sets including lecture video and KITTI data set. In lecture video data set, the speaking content and the notes on presentation slides are extracted from the lecture videos, and dual-variable LDA model involves to cluster the videos. For KITTI data set, derived LDA model is applied to consider continue feature only, and dual-variable LDA model is employed to process two kinds of features. The experimental results show that the proposed method can effectively discover the meaningful semantic characters of the lecture videos.

# Contents

<b>Contents</b>	<b>ix</b>
<b>List of Figures</b>	<b>xi</b>
<b>1 Introduction</b>	<b>1</b>
1.1 Research Challenge . . . . .	2
1.2 Research Objectives . . . . .	3
1.3 Summary of Contributions . . . . .	5
1.4 Outline . . . . .	5
1.4.1 Significance . . . . .	5
1.4.2 Innovations . . . . .	6
1.5 Publications Related to This Thesis . . . . .	7
<b>2 Related works</b>	<b>8</b>
2.1 Background of Trajectory Generation . . . . .	8
2.2 Background of Feature Extraction . . . . .	13
2.3 Background of Trajectory Clustering . . . . .	15
2.3.1 Preliminaries . . . . .	15
2.3.2 Unsupervised Algorithms of Trajectory Clustering . . . . .	21
2.3.3 Supervised Algorithms of Trajectory Clustering . . . . .	29
2.3.4 Semi-supervised Algorithms of Trajectory Clustering . . . . .	34
2.4 Conclusion . . . . .	35
<b>3 Trajectory generation with SIFT</b>	<b>39</b>
3.1 Methodology . . . . .	39

## CONTENTS

3.1.1	Characterize points detection . . . . .	40
3.1.2	Characterize points tracking . . . . .	40
3.2	Experiments . . . . .	43
3.2.1	SURF algorithm . . . . .	44
3.2.2	SIFT algorithm . . . . .	44
3.3	Conclusion . . . . .	46
<b>4</b>	<b>Trajectory Feature Extraction</b>	<b>49</b>
4.1	Methodology . . . . .	50
4.1.1	Continuous Features . . . . .	50
4.1.2	Discrete Features . . . . .	51
4.2	Experiments . . . . .	54
4.3	Conclusion . . . . .	55
<b>5</b>	<b>Dual-variable LDA Model for Trajectory Clustering</b>	<b>57</b>
5.1	Multimodal-LDA Model for Semantic Topic Discovery . . . . .	62
5.1.1	Evidence Extraction . . . . .	63
5.1.2	Multimodal-LDA Model . . . . .	63
5.2	Dual-variable LDA model for Trajectory Clustering . . . . .	73
5.3	Experiments . . . . .	79
5.3.1	Multi-modal LDA model . . . . .	79
5.3.2	Dual-variable LDA model . . . . .	82
5.4	Conclusion . . . . .	85
<b>6</b>	<b>Conclusions</b>	<b>86</b>
<b>A</b>	<b>Results of Video Trajectory Clustering</b>	<b>89</b>
	<b>Bibliography</b>	<b>110</b>

# List of Figures

1.1	The trade-off between features and clustering. . . . .	2
1.2	Algorithm procedure. Objects generate multiple trajectories which start with the characterize point and classify into categories, each category has its unique semantic information representing only one type object. . . . .	6
2.1	Trajectory generated by GPS tracking devices . . . . .	10
2.2	Trajectory generated from camera device . . . . .	10
2.3	Optical Flow algorithm: Optical flow is the pattern of apparent motion of image objects between two consecutive frames caused by the movement of object or camera . . . . .	11
2.4	Optical Flow algorithm in real world application . . . . .	12
2.5	The outlier is separated from normal trajectory. . . . .	13
2.6	The issues need to be fixed in characterizing trajectory data. . . .	14
2.7	For arbitrary trajectory data set, the lengths of trajectories are different from each other. . . . .	16
2.8	DBSCAN for trajectory clustering . . . . .	22
2.9	DBSCAN . . . . .	23
2.10	Hierarchical clustering models . . . . .	25
2.11	Hierarchical clustering models . . . . .	26
2.12	$k$ -NN for trajectory clustering. Inquiry trajectory is the green one, the labeled data are the red and the blue ones which means two clusters. . . . .	30
2.13	CNN is one of the classical models of Neural Network and widely used in images classification . . . . .	37

## LIST OF FIGURES

2.14	Our proposed algorithm clustering trajectories . . . . .	38
3.1	Box filters approximates to Gaussian second order in y- and xy- direction [11]. . . . .	41
3.2	Orientation histogram . . . . .	43
3.3	Results of experiments on the first image of first three sequences in KITTI data set. . . . .	45
3.4	Results of experiments on arbitrary three images of the first se- quence of KITTI data set. . . . .	47
3.5	Results of experiments on arbitrary three images of the second sequence of KITTI data set. . . . .	48
4.1	Top to bottom: 0th, 40th, 80th and 120th frame in 1st sequence of KITTI benchmark. A white van and a cyclist keep staying in the center area of camera image. According to that ego-platform is moving, we have the information that the van and the cyclist is moving. . . . .	52
4.2	Top to bottom: 120th, 130th and 140th frame in 1st sequence of KITTI benchmark. The vehicles parking on the roadside moving a big distance in camera image, such as the silver one moving from center area to border area. . . . .	53
4.3	Results of split camera image into $3 \times 3$ patches. The white van and cyclist are keeping in the center area. . . . .	54
4.4	Results of split camera image into $3 \times 3$ patches. The white van and cyclist are keeping in the center area. . . . .	55
5.1	Multimodel-LDA based topic discovery for lecture videos. . . . .	60
5.2	Dual-variable LDA model for trajectory data. . . . .	61
5.3	(a). standard LDA model. (b). multi-modal LDA model . . . . .	62
5.4	Graphical representation of dual-variable LDA model. . . . .	79