UNIVERSITI PUTRA MALAYSIA

COLLECTIVE INTERACTION FILTERING WITH GRAPH-BASED DESCRIPTORS FOR CROWD BEHAVIOUR ANALYSIS

WONG PEI VOON

FSKTM 2018 85

COLLECTIVE INTERACTION FILTERING WITH GRAPH-BASED DESCRIPTORS FOR CROWD BEHAVIOUR ANALYSIS

By

WONG PEI VOON

Thesis Submitted to the School of Graduate Studies, Universiti Putra Malaysia, in Fulfilment of the Requirements for the Degree of Doctor of Philosophy

December 2018

Abstract of thesis presented to the Senate of Universiti Putra Malaysia in fulfilment of the requirement for the degree of Doctor of Philosophy

COLLECTIVE INTERACTION FILTERING WITH GRAPH-BASED DESCRIPTORS FOR CROWD BEHAVIOUR ANALYSIS

By

WONG PEI VOON

December 2018

Chair : Norwati Mustapha, PhD
Faculty : Computer Science and Information Technology

Crowd behaviour analysis plays an important role in high security interests in public areas such as railway stations, shopping centres, and airports, where large populations gather. Crowd behaviour analysis framework can be divided into low-level, mid-level and high-level. This research is focused on problems of mid-level and high-level. The crowded scenes vary in various densities, structures and occlusion. It brings enormous challenges in effectively dividing detection feature points into cluster to develop dynamic group detector and grouping consistency between frames at mid-level. Besides that, it also poses challenges in identifying generic descriptors to describe motion dynamics caused by pedestrians walk in different directions with extremely diverse behaviours at high-level. Therefore, crowd behaviour analysis framework with enhanced mid and high levels approaches is used in this research to recognise the common properties across different crowded scenes. The recognised common properties are then used to identify generic descriptors from group-level for crowd behaviour classification and crowd video retrieval. At the low-level, motion feature extraction is performed to extract trajectories from each of the video frames. Kanade-Lucas-Tomasi feature point tracker is used to detect and track moving humans, and then tracklets are grouped to form trajectories. At the mid-level, a Collective Interaction Filtering is presented to identify groups by clustering trajectories. It is suitable for group detection in low, medium, and high crowds. At the high-level, the result of Collective Interaction Filtering is used in group motion pattern mining to predict collectiveness, uniformity, stability, and conflict generic descriptors. The generic descriptors identified are represented by graph-based descriptors. Graph-based descriptors are applied to crowd behaviour analysis and crowd video retrieval. All experiments are carried out using CUHK Crowd dataset. The group detection and crowd behaviour analysis ground truth results were provided by related work. The group detection experiment is implemented using the clustering algorithm. Normalized Mutual Information and Rand Index are used to measure the performance of Collective Interaction Filtering. The crowd behaviour analysis experiment is implemented by using non-linear Structural Support Vector Machine with RBF-kernel classifier. Leave-one-out is used to measure the performance of the proposed graph-based descriptors to describe crowd behaviour. The proposed crowd video retrieval approach based on generic descriptors experiment is implemented by using Euclidean distance and Chi-Square distance to measure the similarity matching generic descriptors between the query video and the retrieval set of videos. The crowd video retrieval performance is measured by the average precision in the top k retrieved samples. Experimental results show that the crowd behaviour analysis framework achieves the state-of-the-art performance on the CUHK Crowd dataset. The Collective Interaction Filtering outperforms the related work by achieving 0.55 for Normalized Mutual Information and 0.83 for Rand Index. The average accuracy of the proposed graph-based descriptors for crowd behaviour analysis is 80% compared to the previous works. The proposed crowd video retrieval approach based on graph-based descriptors obtained 49% in average top 10 precision. The performance improvement reveals the effectiveness of the graph-based descriptors for crowd video retrieval in different crowded scenes.

Abstrak tesis yang dikemukakan kepada Senat Universiti Putra Malaysia sebagai memenuhi keperluan untuk ijazah Doktor Falsafah

COLLECTIVE INTERACTION FILTERING DENGAN DESKRIPTOR BERASASKAN GRAF UNTUK ANALISIS TINGKAH LAKU ORANG RAMAI

Oleh

WONG PEI VOON

Disember 2018

Pengerusi : Norwati Mustapha, PhD
Fakulti : Sains Komputer dan Teknologi Maklumat

Analisis tingkah laku orang ramai memainkan peranan penting dalam memastikan keselamatan di kawasan awam seperti stesen keretapi, pusat membeli-belah, dan lapangan terbang, di mana terdapat populasi yang besar berkumpul. Rangka kerja analisis tingkah laku orang ramai boleh dibahagikan kepada tahap rendah, tahap pertengahan dan tahap tinggi. Kajian ini memberi tumpuan kepada permasalahan tahap pertengahan dan tahap tinggi. Adegan yang sesak berbeza mengikut kepelbagaian kepadatan, struktur dan oklusi. Ia membawa cabaran besar dalam membahagikan titik ciri pengesanan yang berkesan ke dalam kluster untuk membangunkan kumpulan pengesan yang dinamik dan pengelompokan yang konsistensi antara bingkai pada peringkat pertengahan. Selain itu, ia juga menimbulkan cabaran dalam mengenal pasti deskriptor generik untuk menggambarkan dinamik pergerakan yang disebabkan oleh pejalan kaki berjalan ke arah yang berbeza dengan tingkah laku yang sangat pelbagai di peringkat tinggi. Oleh itu, rangka kerja analisis tingkah laku orang ramai dengan pendekatan tahap pertengahan dan tinggi yang dipertingkatkan digunakan dalam kajian ini untuk mengenali ciri-ciri umum di seluruh adegan sesak yang berbeza. Ciri umum yang diiktiraf kemudiannya digunakan untuk mengenal pasti deskriptor generik dari peringkat kumpulan untuk klasifikasi kelakuan orang ramai dan pengambilan semula video orang ramai. Di tahap rendah, pengekstrakan ciri gerakan dilakukan untuk mengeluarkan trajektori dari setiap bingkai video. Pengesan titik ciri *Kanade-Lucas-Tomasi* digunakan untuk mengesan dan menjejaki pergerakan manusia, dan kemudiannya dikumpulkan untuk membentuk trajektori. Di tahap pertengahan, *Collective Interaction Filtering* dibentangkan untuk mengenal pasti kumpulan dengan cara pengumpulan trajektori. Teknik ini sesuai digunakkan untuk pengesanan kumpulan dalam kerumunan yang rendah, sederhana mahupun tinggi. Di tahap yang lebih tinggi, hasil *Collective Interaction Filtering* digunakan dalam perlombongan corak gerakan kumpulan untuk meramalkan kolektiviti, keseragaman, kestabilan, dan konflik deskriptor umum. Deskriptor umum yang ditemui diwakili oleh deskriptor berasaskan graf. Deskriptor

berasaskan graf digunakan untuk menganalisis tingkah laku orang ramai dan dapatan video orang ramai. Semua eksperimen dijalankan menggunakan dataset *CUHK Crowd*. Hasil pengesanan kumpulan dan analisis tingkah laku orang ramai telah diperolehi secara langsung berdasarkan kerja-kerja yang telah dilakukan baru-baru ini. Eksperimen pengesanan kumpulan dilaksanakan menggunakan algoritma penggugusan. *Normalized Mutual Information* dan *Rand Index* digunakan untuk mengukur prestasi *Collective Interaction Filtering*. Eksperimen analisis tingkah laku orang ramai dilaksanakan dengan menggunakan mesin vektor sokongan struktur tidak linear bersama pengelas *kernel RBF*. *Leave-one-out* digunakan untuk mengukur prestasi deskriptor berasaskan graf yang dicadangkan untuk menggambarkan kelakuan orang ramai. Pendekatan dapatan video orang ramai yang dicadangkan berdasarkan eksperimen deskriptor umum dilaksanakan dengan menggunakan jarak *Euclidean* dan jarak *Chi-Square* untuk mengukur persamaan deskriptor umum yang sepadan di antara video pertanyaan dan set dapatan video. Prestasi dapatan video orang ramai diukur menggunakan purata ketepatan dalam sampel dapatan video yang teratas. Keputusan eksperimen menunjukkan bahawa rangka kerja analisis tingkah laku orang ramai mencapai prestasi terbaik terhadap dataset *CUHK Crowd*. *Collective Interaction Filtering* melebihi prestasi terbaru yang dicatatkan dengan mencapai 0.55 untuk *Normalized Mutual Information* dan 0.83 untuk *Rand Index*. Purata ketepatan deskriptor berasaskan graf yang dicadangkan bagi analisis tingkah laku orang ramai adalah 80% berbanding dengan kerja sebelumnya. Pendekatan dapatan video orang ramai yang dicadangkan berdasarkan deskriptor berasaskan graf memperoleh purata sebanyak 49% di kalangan 10 ketepatan tertinggi. Peningkatan prestasi menunjukkan keberkesanan deskriptor berasaskan graf bagi tujuan dapatan video orang ramai dalam adegan yang sesak dan berbeza.

ACKNOWLEDGEMENTS

This dissertation would not have been possible without the guidance and the help of several individuals who in one way or another have contributed and extended their valuable assistance in the preparation and completion of this study.

First of all, I am heartily thankful to my supervisor, Associate Professor Dr. Norwati Mustapha, whose encouragement, guidance and support from the initial to the final stage of this thesis enabled me to develop an understanding of the project. I would like to express my sincere thanks and appreciation to the supervisory committee members: Associate Professor Dr. Lilly Suriani Affendey and Associate Professor Dr. Fatimah Khalid for their guidance, valuable suggestions and advice throughout this work in making this research a success. I would also like to express my sincere gratitude to Assistant Professor Dr. Lim Seng Poh, my UTAR colleague for his guidance helped me in writing of this thesis.

My deepest appreciation to my family members, who have been very supportive throughout my study. I also offer my regards and blessings to all of those who supported me in any respect during the completion of the project.

Lastly, thank you to God for guiding me through all the trials I went through in the three years.

This thesis was submitted to the Senate of Universiti Putra Malaysia and has been accepted as fulfilment of the requirement for the degree of Doctor of Philosophy. The members of the Supervisory Committee were as follows:

Norwati Mustapha, PhD
Associate Professor
Faculty of Computer Science and Information Technology
Universiti Putra Malaysia
(Chairman)

Lilly Suriani Affendey, PhD
Associate Professor
Faculty of Computer Science and Information Technology
Universiti Putra Malaysia
(Member)

Fatimah Khalid, PhD
Associate Professor
Faculty of Computer Science and Information Technology
Universiti Putra Malaysia
(Member)

ROBIAH BINTI YUNUS, PhD
Professor and Dean
School of Graduate Studies
Universiti Putra Malaysia

Date:

Declaration by graduate student

I hereby confirm that:

- this thesis is my original work;
- quotations, illustrations and citations have been duly referenced;
- this thesis has not been submitted previously or concurrently for any other degree at any other institutions;
- intellectual property from the thesis and copyright of thesis are fully-owned by Universiti Putra Malaysia, as according to the Universiti Putra Malaysia (Research) Rules 2012;
- written permission must be obtained from supervisor and the office of Deputy Vice-Chancellor (Research and Innovation) before thesis is published (in the form of written, printed or in electronic form) including books, journals, modules, proceedings, popular writings, seminar papers, manuscripts, posters, reports, lecture notes, learning modules or any other materials as stated in the Universiti Putra Malaysia (Research) Rules 2012;
- there is no plagiarism or data falsification/fabrication in the thesis, and scholarly integrity is upheld as according to the Universiti Putra Malaysia (Graduate Studies) Rules 2003 (Revision 2012-2013) and the Universiti Putra Malaysia (Research) Rules 2012. The thesis has undergone plagiarism detection software.

Signature: ________________________ Date: ____________________

Name and Matric No.: Wong Pei Voon, GS42415

Declaration by Members of Supervisory Committee

This is to confirm that:

- the research conducted and the writing of this thesis was under our supervision;
- supervision responsibilities as stated in the Universiti Putra Malaysia (Graduate Studies) Rules 2003 (Revision 2012-2013) are adhered to.

Signature: ______________________

Name of Chairman of Supervisory Committee: Associate Professor Dr. Norwati Mustapha

Signature: ______________________

Name of Member of Supervisory Committee: Associate Professor Dr. Lilly Suriani Affendey

Signature: ______________________

Name of Member of Supervisory Committee: Associate Professor Dr. Fatimah Khalid

TABLE OF CONTENTS

		Page
ABSTRACT		i
ABSTRAK		iii
ACKNOWLEDGEMENTS		v
APPROVAL		vi
DECLARATION		viii
LIST OF TABLES		xii
LIST OF FIGURES		xii
LIST OF ABBREVIATIONS		xvi

CHAPTER

1	**INTRODUCTION**	1
	1.1 Motivation and Background	1
	1.2 Problem Statement	2
	1.3 Research Objectives	3
	1.4 Research Scope	4
	1.5 Research Contributions	4
	1.6 Thesis Outline	5
2	**LITERATURE REVIEW**	7
	2.1 Introduction	7
	2.2 Challenges of Crowd Analysis	10
	2.2.1 Various Appearances	10
	2.2.2 Scales	10
	2.2.3 Various Articulations	11
	2.2.4 Occlusions	11
	2.2.5 View Point Variation	12
	2.2.6 Lighting Changes	13
	2.2.7 Cluttered Background	14
	2.3 Motion Feature Representation	15
	2.3.1 Flow-Based Features	15
	2.3.2 Local Spatio-Temporal Features	16
	2.3.3 Trajectory or Tracklet	17
	2.4 Motion Pattern Segmentation	18
	2.4.1 Segmentation Based on Flow-Based Features	18
	2.4.2 Segmentation Based on Clustering	22
	2.5 Crowd Motion Descriptors	34
	2.5.1 Holistic Approaches	34
	2.5.2 Object-Based Approaches	39
	2.6 Crowd Video Retrieval	51
	2.7 Summary	53
3	**RESEARCH METHODOLOGY**	54
	3.1 Introduction	54
	3.2 Research Methodology Design	54

	3.3	Overview of the CBA Framework	57
	3.4	Evaluation Dataset	63
	3.5	Evaluation Metrics	67
		3.5.1 Metrics of Group Detection	67
		3.5.2 Metrics of Crowd Behaviour Analysis	68
		3.5.3 Metrics of Crowd Video Retrieval	69
	3.6	Summary	70
4	**GROUP DETECTION**		71
	4.1	Introduction	71
	4.2	Group Detection Approach	71
		4.2.1 Create coherent filtering clusters	72
		4.2.2 Determine the key person in every cluster	72
		4.2.3 Compute the weight connectivity of each person with the key person	74
		4.2.4 Infer the people's relationship in a tracklet clusters	75
		4.2.5 Group refinement	75
	4.3	Experimental Results and Evaluation Details	77
	4.4	Summary	84
5	**GENERIC DESCRIPTORS**		85
	5.1	Introduction	85
	5.2	Group Motion Pattern Mining and Prediction	85
		5.2.1 Group Motion Pattern Mining	86
		5.2.2 Kalman Filtering	93
	5.3	Crowd Behaviour Analysis	94
	5.4	Experimental Results and Evaluation Details	94
	5.5	Summary	102
6	**CROWD VIDEO RETRIEVAL**		103
	6.1	Introduction	103
	6.2	Crowd Video Retrieval Approach Based on Generic Descriptors	103
	6.3	Experimental Results and Evaluation Details	104
	6.4	Summary	110
7	**CONCLUSION AND FUTURE WORK**		111
	7.1	Conclusion	111
	7.2	Future Work	112

REFERENCES	114
BIODATA OF STUDENT	122
LIST OF PUBLICATIONS	123

LIST OF TABLES

Table		Page
2.1	Summary of approaches in motion pattern segmentation	33
2.2	Summary of approaches in crowd behaviour analysis	50
3.1	Comparison details of group detection approach	60
3.2	Quantitative comparison of step analysis	61
3.3	Comparison details of crowd behaviour analysis based on crowd features or descriptors	61
3.4	Comparison details of linear and nonlinear SVMs to classify crowd video clips	62
3.5	Graph-based descriptors evaluation	62
3.6	Comparison details of crowd video retrieval approach based on generic descriptors and related crowd features	62
3.7	Crowd density ranges	64
5.1	List of group size calculation	87
5.2	Generic Descriptors	91
5.3	List of crowd behaviour classes	94
5.4	Comparison for Linear and Nonlinear SVMs	100
6.1	Crowd video retrieval attributes	103

LIST OF FIGURES

Table		Page
2.1	Structured Crowded Scenes	7
2.2	Unstructured Scenes	8
2.3	Schematic Representation of the Topics in Crowd analysis Field	9
2.4	Examples of Various Appearances	10
2.5	Wide Range of People Scales	11
2.6	Human Body could Show Various Postures	11
2.7	Inter-Object Occlusion	12
2.8	Scene Occlusion	12
2.9	View Point Variation	13
2.10	Day Time Caused Lighting Changes	13
2.11	Various Event Caused Lighting Changes	14
2.12	Cluttered Background	14
2.13	(a) The Optic Flow Experienced by a Rotating Observer (b) Optical Flow (c) Particle Flow (d) Streak Flow	15
2.14	Spatio- temporal Subvolumes and Hidden Markov Model Train	16
2.15	Motion Histograms of the Three Pixels with Different Moving Directions	17
2.16	Example of Tracklets	18
2.17	The Flow Segmentations for a Number of Sequences	19
2.18	Streak Flow	20
2.19	Crowd Segmentation Algorithm	20
2.20	Architecture of Segmentation Method based on Streakline	21
2.21	Samples from GaTech Football Play Dataset	24
2.22	Algorithm CoheFilterDet of Coherent Filtering	25
2.23	Motion Pattern Segmentation Framework	26
2.24	a) Dynamic Texture b) Dynamic Texture Mixture. The Hidden Variable z Selects the Parameters of the Remaining Nodes	27
2.25	Flowchart of Unsupervised Approach for Video Scene Understanding	28
2.26	Process Flow of Scene Understanding by Statistical Modeling of Motion Patterns	28
2.27	The Factor Graph for the Third-order CRF Model	29
2.28	(A) Graphical Representation of the RFT Model. (B) Example of RFT Model	30
2.29	Flow Chart of Two-Level Motion Pattern Mining	30
2.30	The Algorithm Learning the CT Priors for Group Discovery	31
2.31	Physical Identity	32
2.32	Social Identity	32
2.33	(a) An Illustration of Streaklines for a Video Sequence. (b) The Crowd Segmentation Algorithm. (c) Abnormal Behaviour Detection Algorithm	35
2.34	The Computation of Divergence Factor	35
2.35	Framework for Anomaly Detection and Localization	36

2.36	Motion Pattern Extraction and Event Detection Algorithm Steps	37
2.37	Five Crowd Behaviours	37
2.38	A Single HMM for Each Tube of Observations	38
2.39	Framework of Crowd Behaviour Perception Algorithm	39
2.40	Anomaly Detection using Social Force Model Algorithm	39
2.41	The Process of using the BOW Method to Describe a Clip	41
2.42	Mixture Model of Dynamic Pedestrian-Agents (MDA)	42
2.43	Diagram Illustrating the Overall Flow of Lagrangian Eigenmap Method	43
2.44	Spatio-Temporal Compositions	43
2.45	Flow Complexity Computation Algorithm	44
2.46	Main Steps for Crowd Escape Behaviour Detection	45
2.47	Learning Scene-Independent Group Descriptors for Crowd Understanding Framework	46
2.48	The Deep Learning Model	46
2.49	The Flow Diagram of Framework based on Trajectory and Pixel-based Features	47
2.50	Trajectory-based Motion Coding Algorithm for Extracting the CDT Descriptors	48
2.51	Bilinear CD Descriptors Algorithm	49
2.52	Crowd Video Retrieval based on Deep Attribute-embedding Graph Ranking Framework	52
3.1	Research Methodology Steps	55
3.2	CBA Framework	58
3.3	Trajectories	59
3.4	KLT Feature Point Tracker Results	59
3.5	Tracklets People Walking in a Shopping Mall	60
3.6	Video Clips of the Different Video Resolutions	63
3.7	Detail Information of Dataset	65
3.8	Sample for Selected Video Clips of the Different Crowded Scenes	66
3.9	RI Resulting in Contingency Table	67
3.10	Leave-one-out	68
3.11	Example of Confusion Matrix for Leave-one-out Evaluation	68
3.12	Summary Outcomes of the Classification in a Confusion Matrix	69
4.1	Example for CI	72
4.2	Example First Cluster Trajectories from Frame 1 to Frame 3	73
4.3	Influence of Video Frames to CI Approach Evaluation	77
4.4	The Quantitative Comparison of Group Detection Methods	78
4.5	Examples of Ground Truth, CF, CT, and CI Group Detection Results for Indoor Crowded Scenes	79
4.6	Examples of Ground Truth, CF, CT, and CI Group Detection Results for Outdoor Crowded Scenes	80
4.7	Quantitative Comparison of CF, CT, CI Methods in Different Crowd Densities	81
4.8	Quantitative Comparison of CF, CT, CI Methods in Different Crowded Scenes in NMI	82
4.9	Quantitative Comparison of CF, CT, CI Methods in Different Crowded Scenes in RI	82

4.10	Quantitative Comparison of Step Analysis	83
5.1	The Proposed Group Motion Pattern Mining and Prediction Approach	86
5.2	Group Size Example	87
5.3	Examples for Evenly Space	89
5.4	Example of Speed with Different Direction Location in Groups	90
5.5	Example for Generic Descriptors	91
5.6	Kalman Filtering Steps	93
5.7	Example for 8 Classes Crowd Behaviour	95
5.8	Confusion Matrices of Crowd Video Classification (darker colour represents higher accuracy)	97
5.9	The Quantitative Average Accuracy Comparison of the Proposed Graph-based Descriptors with Previous Works	98
5.10	Example for Non-linear SVM	101
5.11	Quantitative Comparison of Components in the Proposed Graph-based Descriptors	102
6.1	Example of Similarity between Videos based on Generic Descriptors	103
6.2	Example Top 5 Crowd Video Retrieval Result	105
6.3	The Quantitative Average Precision at Top k Comparison of the Graph-based Descriptors with Previous Works	106
6.4	The Quantitative Average Precision at Top 100 of each Crowd Behaviour Classes	108
6.5	The Example of Query Videos in 1, 3 and 5 Crowd Behaviour Classes and Their Retrieval Results	109

LIST OF ABBREVIATIONS

CBA	Crowd Behaviour Analysis
CUHK	Chinese University of Hong Kong
CI	Collective Interaction
EM	Expectation-Maximization
RBF	Radial Basis Function
NMI	Normalized Mutual Information
RI	Rand Index
CF	Coherent Filtering
CT	Collective Transition
SVM	Structural Support Vector Machine
X^2	Chi-Square distance
HMMs	Hidden Markov Models
LCS	Lagrangian Coherent Structures
LCSS	Longest Common Subsequence
MHI	Motion History Image
MHOF	Multi-scale Histogram of Optical Flow
2-D	Two-dimensional
3-D	Three-dimensional
LTDS	Local–translational Domain Segmentation
ODFs	Orientation Distribution Functions
CRF	Conditional Random Field
RFT	Random Field Topic
MRF	Markov Random Field
LDA	Latent Dirichlet Allocation
KLT	Kanade–Lucas–Tomasi
K-NN	K-Nearest-Neighbor
DFS	Depth-First Search
PDFs	Probability Density Functions
MDT	Mixture of Dynamic Texture
BOW	Bag-of-words
SRC	Sparse Reconstruction Cost
LAE	Local Abnormal Event
GAE	Global Abnormal Event
MDA	Mixture Model of Dynamic Pedestrian-Agents
NMC	Neighborhood Motion Concurrence
CD	Curl and divergence
DC	Distance Connectivity
OC	Occurrence Connectivity
SC	Speed Connectivity
GS	Group Size
ES	Evenly Space
SD	Speed Direction
TP	True Positive
TN	True Negative
FP	False Positive
FN	False Negative

AP@k	Average precision in the top k
GMM	Gaussian mixture model

CHAPTER 1

INTRODUCTION

1.1 Motivation and Background

Crowd is the agglomeration of many people with different kinds of behaviour in the same area at the same time (Junior et al., 2010) such as music festivals, sports events, railway stations, airports, shopping mall and other such places. In general, the security of large gatherings is the most important feature. Any abnormal behaviour or incident in dense crowds would cause undesirable happenings because of the synergic result of human relations (Mehran et al., 2009; Mehran et al., 2010; Murino et al., 2017). As the size of the crowd becomes larger, the harder it becomes to monitor their actions with the human eye (Rodriguez et al., 2017).

Researchers have turned to surveillance technology based on computer vision to monitor crowds automatically in helping to discover crowd disasters. Computer vision surveillance technology is also used for people management such as tourist flow estimation or pedestrian traffic management in recent years. However, dangerous and criminal behaviours are mostly observed within groups of people (Jacques et al., 2007; Mora Colque et al., 2014; Solera et al., 2016; Shao et al., 2017). Therefore, research surveillance community changed from the monitoring of a single person or a crowd population to group of people and their behaviour (Jacques et al., 2007; Mora Colque et al., 2014; Solera et al., 2016; Shao et al., 2017; Wang et al., 2017). Group is defined as a collection of individuals assembled together in the same place who interact with one another, share similar characteristics, and collectively have a sense of unity (Ge et al., 2012). The crowd is mainly composed of groups rather than individuals, so focusing on the group helps to understand the crowd, and vice versa (Murino et al., 2017). Crowd behaviour in different crowded scenes can be analysed based on group descriptors (Shao et al., 2017).

From a computer vision point of view, the study of group or crowd analysis for crowded scenes are generally modelled after a three-level approach (Murino et al., 2017). At the low-level, moving objects are discovered and tracked to extract the crowd motion features from each of the video frames. Motion features such as particle flow (Ali & Shah, 2007), streak flow (Mehran et al., 2010), spatio-temporal (Kratz & Nishino, 2009, 2012), and trajectory or tracklet (Zhou et al., 2011, 2012; Shao et al., 2017) are extracted. At the mid-level, motion pattern segmentation is used in crowd analysis by grouping the features into similar categories through some resemblance measures or probabilities (Li et al., 2015). At the high-level, a semantic understanding of the group or crowd behaviour is obtained.

However, most studies on crowd behaviour analysis focus on specific scenes resulting in model overfitting, and thus are hardly useful for other scenes (Mehran et al., 2009; Junior et al., 2010; Thida et al., 2013; Li et al., 2015; Kok et al., 2016; Shao et al., 2017; Wang & Loy, 2017).

1.2 Problem Statement

The study of group or crowd analysis for crowded scenes are generally modelled in low, middle and high levels (Murino et al., 2017). Low-level algorithms have been widely studied in the field of computer vision, and achieved gratifying results. However, algorithms at the middle and high level have been just started studied in recent times. The motion features are the most useful and representative part in video frames which can help in describing crowd behaviour. Therefore, trajectory or tracklet motion feature applied in this research at the low-level. Trajectory or tracklet motion feature researches achieved relatively better performance for structured and unstructured crowded scenes (Zhou et al., 2014; Solera et al., 2016; Shao et al., 2017; Wang et al., 2017). This research aims to tackle the following challenges from the middle and high levels for different crowded scenes.

Tracking results are usually described as trajectory, and short-range trajectory fragments are defined as tracklets (Zhou et al., 2011). However, occlusion affects the terminations of tracklets cause complete trajectories are hard to gain (Zhou et al., 2011; Zhou et al., 2014; Solera et al., 2016; Shao et al., 2017; Wang et al., 2017). For that reason, most existing studies at mid-level follow holistic approach in which the crowd is considered as a single entity to segment the motions (Ali & Shah, 2007; Mehran et al., 2009; Mehran et al., 2010; Li et al., 2015; Wu et al., 2017). The holistic approach focuses on crowd behaviour identification as a whole. Yet, holistic approach is only suitable to be applied in high-density population of structured crowded scenes. Hence, group detection approaches which can identify groups by clustering trajectories are proposed. Group detection approaches provide a trade-off between the holistic approach and the individual-based approach for crowd behaviour analysis (Ge et al., 2012; Zhou et al., 2012; Wang et al., 2013; Solera et al., 2016; Shao et al., 2017). Current group detection approaches suffer from the following problems: 1) Current clustering techniques cannot effectively divide the detection feature points into cluster to develop dynamic of group detector. The detection feature points across frames can be lost due to occlusion (Liang et al., 2014; Solera et al., 2016; Trojanová et al., 2016; Shao et al., 2017; Wang et al., 2017). 2) Crowds with various densities and structures (Junior et al., 2010; Li et al., 2015; Kok et al., 2016; Shao et al., 2017). 3) Many previous works (Zhou et al., 2012; Zhou et al., 2014; Trojanová et al., 2016; Shao et al., 2017) emphasis on the motion correlation of persons within a local area and limited to grouping consistency between frames.

The high-level focuses on discovery of crowd motion descriptors based on low-level motion features in order to facilitate understanding crowd behaviour. Therefore, various crowd motion descriptors such as social force (Mehran et al., 2009), potential field (Mehran et al., 2010), chaotic invariants (Wu et al., 2010), spatio-temporal gradients (Kratz & Nishino, 2009, 2012), eigenvalues (Solmaz et al., 2012), spatio-temporal viscous fluid field (Su et al., 2013), multi-scale histogram of optical flow (Cong et al., 2013) and collectiveness (Zhou et al., 2014) have been suggested from different perspectives for scene-specific crowd behaviour analysis. These crowd motion descriptors are restricted to holistic perspective. Recently, some researchers proposed generic descriptors, such as curl, divergence, collectiveness, uniformity, stability, and conflict, from the computer vision point of view to describe crowd behaviour for different crowded scenes (Shao et al., 2015; Shao et al., 2017; Wang & Loy, 2017; Wu et al., 2017; Wu et al., 2017). However, these generic descriptors cannot perform well for the motion dynamics caused by pedestrians who walk in different directions with extremely diverse behaviours; such as pedestrians in streets or shopping malls. Current tracking approaches are difficult in capturing accurately motion interaction among people in different crowded scenes. Crowds with various densities, structures and occlusion affect the efficiency of generic description in classifying crowd behaviour for different crowded scenes accurately (Shao et al., 2015; Shao et al., 2017; Wang & Loy, 2017; Wu et al., 2017; Yi et al., 2017).

In recent years, researchers have shift their research interest to retrieve the preferred videos based on measuring the likeness between video queries and crowd patterns contained in crowd videos (Zhang et al., 2016; Shao et al., 2017; Wu et al., 2017). Crowd videos are difficult to segment the motion pattern because of people move to occlude each other or blocked by non-human items (Liang et al., 2014; Solera et al., 2016; Trojanová et al., 2016; Zhang et al., 2016; Shao et al., 2017; Wang et al., 2017). Besides that, it is also challenging to segment the motion pattern of crowds with low, medium and high densities in structured and unstructured crowd scenes (Junior et al., 2010; Li et al., 2015; Kok et al., 2016; Shao et al., 2017). These characteristics cause difficulty in identifying generic descriptors to describe crowd patterns (Shao et al., 2015; Shao et al., 2017; Wang & Loy, 2017; Wu et al., 2017; Yi et al., 2017), which commonly cause the difficulty in measuring the likeness between video queries and crowd patterns enclosed in crowd videos (Zhang et al., 2016; Shao et al., 2017; Wu et al., 2017).

In summary, the main limitation of the above discussed approaches includes difficulty in object detection and tracking when occlusion occurs in scenes which affects the accuracy of group clustering, crowd behaviour classification and crowd video retrieval.

1.3 Research Objectives

The main aim of this research is development of crowd behaviour analysis (CBA) framework with enhanced mid and high levels approaches to recognise the common properties across different crowded scenes. The recognised common properties are then used to identify generic descriptors from group-level for crowd behaviour classification and crowd video retrieval. The accuracy of this framework will be evaluated through extensive experiments. To achieve the objective, the following ideas are adopted:

- To propose a group detection approach with abilities to accurately identify groups by clustering trajectories in crowds with various densities, structures and occlusion of each other. It also tackles grouping consistency between frames.
- To propose group motion pattern mining and prediction approach to identify generic descriptor for crowd behaviour classification.
- To propose a crowd video retrieval approach based on generic descriptors.

1.4 Research Scope

The scope of this research is focused on the crowd behaviour analysis application offline processing. The more advanced case of real time processing is left for the future work since the implementation of real-time application normally requires hardware with high-specs. This research is concerned with crowd behaviour analysis for human interaction occurring in indoor and outdoor video surveillance such as streets, shopping malls and stations with low, medium and high population densities. All experiments are carried out using CUHK Crowd Dataset (Shao et al., 2017). Trajectories are the motion feature extracted when an object is moving. This experiments assumes that all the trajectories extracted are from the moving humans. Besides that, crowd behaviour classification and crowd video retrieval are only based on collectiveness, uniformity, stability, and conflict generic descriptors.

1.5 Research Contributions

The core contribution of this research is proposal of an enhanced approach for CBA framework to identify generic descriptors from group-level for crowd behaviour classification and crowd video retrieval. The framework has motion feature extraction, group detection, generic descriptors, and crowd video retrieval components. Each component except motion feature extraction has their own contributions as follows:

- A group detection approach is proposed. The first contribution is the ability to determine the key person which remains consistent between all frames in each cluster over time-varying dynamics in crowded scenes to handle grouping consistency between frames. The second contribution is to form an inference about human relationships using Expectation-Maximization (EM) algorithm based on distance, occurrence, and speed correlations of each person with the key person to handle the occlusion. The final contribution is a group refinement threshold based on the results gathered through the inferences on human relationships in order to tackle the crowds with various densities and structures.
- A group motion pattern mining and prediction approach is proposed to identify the accuracy of collectiveness, uniformity, stability, and conflict generic descriptors for behaviour understating in different crowded scenes. A graph partitioning algorithm with assigned group size, evenly space, and speed direction connection between pairwise members in group weights to

each edge of undirected graph to mining the group interaction pattern. Kalman filtering is applied in generic descriptor prediction in order to obtain a precise prediction of their motion interaction over the frame in different crowded scenes and to tackle the problem of occlusion.

- An effective crowd video retrieval approach that employs collectiveness, uniformity, stability, and conflict generic descriptors. The Euclidean distance and Chi-Square (X^2) distance are used to measure the similarity matching between the query video and the remaining video clips.

1.6 Thesis Outline

Several approaches will be presented to address the critical problems of accurate object detection and tracking when occlusion occurs in scenes, which will affect the efficiency of group detection, crowd behaviour classification, and crowd video retrieval. To achieve the overall goal, each chapter shall describe a component of this research and each chapter is arranged as follows:

Chapter 1: Introduction. The introduction chapter provides a brief background of the research work. Then, it addresses the main challenges of important tasks in automation of crowd behaviour analysis. This chapter also highlights the objectives, scopes and contributions of this research. Finally, the thesis outline is provided, offering a summary of each chapter.

Chapter 2: Literature Review. This chapter discusses the review of current researches for motion feature extraction, motion pattern segmentation and crowd motion descriptors.

Chapter 3: Research Methodology. This chapter presents the steps taken for this research and the research methodology employed. The CBA framework with enhanced mid and high levels approaches is used as a solution to the limitations of current crowd behaviour analysis approaches that have been found though literature review. The evaluation dataset and metrics used in this research work are also explained in this chapter.

Chapter 4: Group Detection. This chapter presents the proposed group detection approach to identify groups by clustering trajectories. The accuracy of the proposed approach is evaluated and compared with related works through a set of experiments.

Chapter 5: Generic Descriptors. This chapter presents the group motion pattern mining and prediction approach to identify the accuracy of collectiveness, uniformity, stability, and conflict generic descriptors for group behaviour understating in different crowded scenes. Then, the non-linear Structural Support Vector Machine (SVM) with RBF-kernel classifier is trained based on the prediction generic descriptors results in order to distinguish among the crowd videos. The results of extensive experiments are described to show its effectiveness over other crowd motion descriptors.

Chapter 6: Crowd Video Retrieval. This chapter presents the proposed crowd video retrieval approach based on generic descriptors. This chapter also demonstrates the evaluation results of the proposed approach.

Chapter 7: Conclusion and Future Work. This chapter offers concise conclusions on the outlined objective and highlights some recommendation for future works.

REFERENCES

Aimonen, P. (2011). Basic concept of Kalman filtering. https://commons.wikimedia.org/wiki/File:Basic_concept_of_Kalman_filtering.svg Retrieved 28 March 2017.

Ali, S. (2013). Measuring flow complexity in videos. In *Proceedings of the IEEE International Conference on Computer Vision*, (pp.1097–1104).

Ali, S., & Shah, M. (2007). A Lagrangian particle dynamics approach for crowd flow segmentation and stability analysis. In *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, (pp.1–6).

Ali, S., & Shah, M. (2013). Crowd Flow Segmentation using Lagrangian Particle Dynamics, *Modeling, Simulation and Visual Analysis of Crowds: A Multidisciplinary Perspective* pp. 213-236. New York, NY: Springer New York.

Araujo, A., & Girod, B. (2018). Large-Scale Video Retrieval Using Image Queries. *IEEE Transactions on Circuits and Systems for Video Technology*, 28(6), 1406–1420.

Bagherpour, P., Cheraghi, S. A., & Mokji, M. bin M. (2012). Upper Body Tracking Using KLT and Kalman Filter. *Procedia Computer Science*, *13*, 185–191.

Benabbas, Y., Ihaddadene, N., & Djeraba, C. (2011). Motion Pattern Extraction and Event Detection for Automatic Visual Surveillance. *EURASIP Journal on Image and Video Processing*, 1–15.

Brox, T., Papenberg, N., & Weickert, J. (2004). High Accuracy Optical Flow Estimation Based on a Theory for Warping. *Computer Vision - ECCV 2004*, *4*(May), 25–36.

Chamasemani, F. F., Affendey, L. S., Mustapha, N., & Khalid, F. (2015). A Framework for Automatic Video Surveillance Indexing and Retrieval. *Research Journal of Applied Sciences, Engineering and Technology*, *10*(11), 1316–1321.

Chamasemani, F. F., Affendey, L. S., Mustapha, N., & Khalid, F. (2017). Speeded up surveillance video indexing and retrieval using abstraction. In *Proceedings of the IEEE International Conference on Signal and Image Processing Applications (ICSIPA)* (pp. 374–378).

Chan, A. B., & Vasconcelos, N. (2008). Modeling, Clustering, and Segmenting Video with Mixtures of Dynamic Textures. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *30*(5), 909–926.

Chang, C., & Lin, C. (2011). LIBSVM. *ACM Transactions on Intelligent Systems and Technology*, *2*(3), 1–27.

Cheng, Z., Shen, J., & Miao, H. (2016). The effects of multiple query evidences on social image retrieval. *Multimedia Systems*, *22*(4), 509–523.

Cheriyadat, A. M., & Radke, R. J. (2008). Detecting dominant motions in dense crowds. *IEEE Journal on Selected Topics in Signal Processing*, *2*(4), 568–581.

Chlis, N. K. (2013). *Comparison of Statistical Methods for Genomic Signature Extraction*, Diploma Thesis, Technical University of Crete.

Christopher D. Manning, P. R., & Schütze, H. (2009). *Introduction to Information Retrieval*. *Cambridge University Press* (Vol. 35).

Cong, Y., Yuan, J., & Liu, J. (2013). Abnormal event detection in crowded scenes using sparse representation. *Pattern Recognition*, *46*(7), 1851–1864.

Cosar, S., Donatiello, G., Bogorny, V., Garate, C., Alvares, L. O., & Bremond, F. (2017). Toward Abnormal Trajectory and Event Detection in Video Surveillance. *IEEE Transactions on Circuits and Systems for Video Technology*, *27*(3), 683–695.

Dempster, A. P., Laird, N. M., & Rubin, D. B. (1977). Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society Series B Methodological*, *39*(1), 1–38.

Deza, E., & Deza, M. M. (2009). *Encyclopedia of Distances*. *Encyclopedia of Distances*. Berlin, Heidelberg: Springer Berlin Heidelberg.

Dupont, C., Tobias, L., & Luvison, B. (2017). Crowd-11: A Dataset for Fine Grained Crowd Behaviour Analysis. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)* (pp. 2184–2191).

Ge, W., Collins, R. T., & Ruback, R. B. (2012). Vision-Based Analysis of Small Groups in Pedestrian Crowds. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *34*(5), 1003–1016.

Ghule, R. S., & M. Shinde, S. (2014). A Survey Paper on Trajectory Pattern Mining for Pattern Matching Query. *International Journal of Computer Applications*, *86*(17), 27–30.

Giannotti, F., Nanni, M., Pinelli, F., & Pedreschi, D. (2007). Trajectory pattern mining. In *Proceedings of the 13th ACM SIGKDD international conference on Knowledge discovery and data mining - KDD '07* (pp. 330-340).

Gupta, M. R. (2010). Theory and Use of the EM Algorithm. *Foundations and Trends® in Signal Processing*, *4*(3), 223–296.

Haritaoglu, I., & Flickner, M. (2001). Detection and tracking of shopping groups in stores. In *Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. CVPR 2001* (Vol. 1, p. I-431-I-438).

Hu, M., Ali, S., & Shah, M. (2008a). Detecting global motion patterns in complex videos. In *Proceedings of the 2008 19th International Conference on Pattern Recognition* (pp. 1–5).

Hu, M., Ali, S., & Shah, M. (2008b). Learning motion patterns in crowded scenes using motion flow field. In *Proceedings of the 2008 19th International Conference on Pattern Recognition* (pp. 1–5).

Hu, W., Tan, T., Wang, L., & Maybank, S. (2004). A Survey on Visual Surveillance of Object Motion and Behaviors. *IEEE Transactions on Systems, Man and Cybernetics, Part C (Applications and Reviews)*, *34*(3), 334–352.

Hu, W., Xiao, X., Fu, Z., Xie, D., Tan, T., & Maybank, S. (2006). A system for learning statistical motion patterns. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *28*(9), 1450–1464.

Hu, W., Xie, D., Fu, Z., Zeng, W., & Maybank, S. (2007). Semantic-Based Surveillance Video Retrieval. *IEEE Transactions on Image Processing*, *16*(4), 1168–1181.

Hu, W., Xie, D., Fu, Z., Zeng, W., & Maybank, S. (2011). A Survey on Visual Content-Based Video Indexing and Retrieval. *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*, *41*(6), 797–819.

Idrees, H., Warner, N., & Shah, M. (2014). Tracking in dense crowds using prominence and neighborhood motion concurrence. *Image and Vision Computing*, *32*(1), 14–26.

Jacques, J. C. S., Braun, A., Soldera, J., Musse, S. R., & Jung, C. R. (2007). Understanding people motion in video sequences using Voronoi diagrams. *Pattern Analysis and Applications*, *10*(4), 321–332.

Jalali, M., Mustapha, N., Mamat, A., & Sulaiman, M. N. B. (2008). A new clustering approach based on graph partitioning for navigation patterns mining. In *Proceedings of the 2008 19th International Conference on Pattern Recognition* (Vol. 4, pp. 1–4).

Jiang, M., Huang, J., Wang, X., Tang, J., & Wu, C. (2014). An Approach for Crowd Density and Crowd Size Estimation. *Journal of Software*, *9*(3), 757–762.

Jodoin, P. M., Benezeth, Y., & Wang, Y. (2013). Meta-tracking for video scene understanding. In *Proceedings of the 2013 10th IEEE International Conference on Advanced Video and Signal Based Surveillance, AVSS 2013* (pp. 1–6).

Junior, J. C. S. J., Musse, S. R., & Jung, C. R. (2010). Crowd Analysis Using Computer Vision Techniques. *IEEE Signal Processing Magazine*, *27*(September), 66–77.

Khokher, M. R., Bouzerdoum, A., & Phung, S. L. (2014). Crowd Behavior Recognition Using Dense Trajectories. In *Proceedings of the 2014 International Conference on Digital Image Computing: Techniques and Applications (DICTA)* (pp. 1–7).

Kok, V. J., Lim, M. K., & Chan, C. S. (2016). Crowd behavior analysis: A review where physics meets biology. *Neurocomputing*, *177*, 342–362.

Kratz, L., & Nishino, K. (2009). Anomaly detection in extremely crowded scenes using spatio-temporal motion pattern models. In *Proceedings of the 2009 IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops, CVPR Workshops 2009*, 1446–1453.

Kratz, L., & Nishino, K. (2012). Tracking pedestrians using local spatio-temporal motion patterns in extremely crowded scenes. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *34*(5), 987–1002.

Lafon, S., & Lee, A. B. (2006). Diffusion maps and coarse-graining: A unified framework for dimensionality reduction, graph partitioning, and data set parameterization. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *28*(9), 1393–1403.

Lee, J.-G., Han, J., & Li, X. (2015). A Unifying Framework of Mining Trajectory Patterns of Various Temporal Tightness. *IEEE Transactions on Knowledge and Data Engineering*, *27*(6), 1478–1490.

Levenshtein, V. I. (1966). Binary codes capable of correcting deletions, insertions, and reversals. *Soviet Physics Doklady*.

Li, F., & Niebles, J. C. (2016). Basic Concepts of Segmentation. *http://vision.stanford.edu/teaching/cs131_fall1617/lectures/lecture12_clustering_and_seg_cs131_2016.pdf* Retrieved 23 October 2016.

Li, F., & Niebles, J. C. (2016). K-means Clustering. *http://vision.stanford.edu/teaching/cs131_fall1617/lectures/lecture13_kmeans_mean_shift_cs131_2016* Retrieved 23 October 2016.

Li, F., & Niebles, J. C. (2016). Linear Classifiers & Classification. *http://vision.stanford.edu/teaching/cs131_fall1617/lectures/lecture15_linear_classifiers_and_classification_cs131_2016.pdf* Retrieved 6 November 2016.

Li, F., & Niebles, J. C. (2016). Basic Concepts in Recognition, Nearest Neighbor Match. *http://vision.stanford.edu/teaching/cs131_fall1617/lectures/lecture17_intro_objrecog_cs131.pdf* Retrieved 17 February 2017.

Li, R., & Chellappa, R. (2010). Group motion segmentation using a Spatio-Temporal Driving Force Model. In *Proceedings of the 2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition* (pp. 2038–2045).

Li, R., Chellappa, R., & Zhou, S. K. (2013). Recognizing Interactive Group Activities Using Temporal Interaction Matrices and Their Riemannian Statistics. *International Journal of Computer Vision*, *101*(2), 305–328.

Li, T., Chang, H., Wang, M., Ni, B., Hong, R., & Yan, S. (2015). Crowded Scene Analysis: A Survey. *IEEE Transactions on Circuits and Systems for Video Technology*, *25*(3), 367–386.

Liang, R., Zhu, Y., & Wang, H. (2014). Counting crowd flow based on feature points. *Neurocomputing*, *133*, 377–384.

Mehran, R., Moore, B. E., & Shah, M. (2010). A streakline representation of flow in crowded scenes. *LNCS*, *6313 LNCS*(PART 3), 439–452.

Mehran, R., Oyama, A., & Shah, M. (2009). Abnormal crowd behavior detection using social force model. In *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops, CVPR Workshops 2009* (pp. 935–942).

Mora Colque, R. V. H., Camara Chavez, G., & Schwartz, W. R. (2014). Detection of Groups of People in Surveillance Videos Based on Spatio-Temporal Clues. In *Proceedings of the Pattern Recognition, Image Analysis, Computer Vision, and Applications* (Vol. 8827, pp. 948–955).

Murino, V., Cristani, M., Shah, S., & Savarese, S. (2017). *Chapter 1 - The Group and Crowd Analysis Interdisciplinary Challenge*. *Group and Crowd Behavior for Computer Vision* (1st ed.). Elsevier Inc.

Nielsen, F. (2012). K-MLE: A fast algorithm for learning statistical mixture models. In *Proceedings of the 2012 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)* (pp. 869–872).

Pele, O., & Werman, M. (2010). The Quadratic-Chi Histogram Distance Family. *LNCS* (Vol. 6312 LNCS, pp. 749–762).

Pellegrini, S., Ess, A., & Van Gool, L. (2010). Improving Data Association by Joint Modeling of Pedestrian Trajectories and Groupings. *LNCS* (Vol. 6311 LNCS, pp. 452–465).

Rodriguez, M., Ali, S., & Kanade, T. (2009). Tracking in unstructured crowded scenes. In *Proceedings of the IEEE International Conference on Computer Vision* (pp. 1389–1396).

Rodriguez, M., Sivic, J., & Laptev, I. (2017). The Analysis of High Density Crowds in Videos. In *Group and Crowd Behavior for Computer Vision* (1st ed., pp. 89–113). Elsevier.

Rodriguez, M., Sivic, J., Laptev, I., & Audibert, J. Y. (2011). Data-driven crowd analysis in videos. In *Proceedings of the IEEE International Conference on Computer Vision* (pp. 1235–1242).

Roshtkhari, J. M., & Levine, M. D. (2013). An on-line, real-time learning method for detecting anomalies in videos using spatio-temporal compositions. *Computer Vision and Image Understanding*, *117*(10), 1436–1452.

Saleemi, I., Hartung, L., & Shah, M. (2010). Scene Understanding by Statistical Modeling of Motion Patterns. In *Proceedings of the IEEE Conference On Computer Vision and Pattern Recognition (CVPR),* 2069–2076.

Shadden, S. C., Lekien, F., & Marsden, J. E. (2005). Definition and properties of Lagrangian coherent structures from finite-time Lyapunov exponents in two-dimensional aperiodic flows. *Physica D: Nonlinear Phenomena*, *212*(3–4), 271–304.

Shao, J., Kang, K., Loy, C. C., & Wang, X. (2015). Deeply learned attributes for crowded scene understanding. In *Proceedings of the 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (pp. 4657–4666).

Shao, J., Loy, C. C., & Wang, X. (2017). Learning Scene-Independent Group Descriptors for Crowd Understanding. *IEEE Transactions on Circuits and Systems for Video Technology*, *27*(6), 1290–1303.

Solera, F., Calderara, S., & Cucchiara, R. (2016). Socially Constrained Structural Learning for Groups Detection in Crowd. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *38*(5), 995–1008.

Solmaz, B., Moore, B. E., & Shah, M. (2012). Identifying Behaviors in Crowd Scenes Using Stability Analysis for Dynamical Systems. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *34*(10), 2064–2070.

Song, L., Jiang, F., Shi, Z., Molina, R., & Katsaggelos, A. K. (2014). Toward Dynamic Scene Understanding by Hierarchical Motion Pattern Mining. *IEEE Transactions on Intelligent Transportation Systems*, *15*(3), 1273–1285.

Su, H., Yang, H., Zheng, S., Fan, Y., & Wei, S. (2013). The Large-Scale Crowd Behavior Perception Based on Spatio-Temporal Viscous Fluid Field. *IEEE Transactions on Information Forensics and Security*, *8*(10), 1575–1589.

Sze, V., Chen, Y., Yang, T., & Emer, J. S. (2017). Efficient Processing of Deep Neural Networks: A Tutorial and Survey. *Proceedings of the IEEE*, 105(12), 2295–2329.

Thida, M., Eng, H., & Remagnino, P. (2013). Laplacian Eigenmap With Temporal Constraints for Local Abnormality Detection in Crowded Scenes. *IEEE Transactions on Cybernetics*, *43*(6), 2147–2156.

Thida, M., Yong, Y. L., Climent-pérez, P., Eng, H., & Remagnino, P. (2013). A Literature Review on Video Analytics of Crowded Scenes. In *Intelligent Multimedia Surveillance* (pp. 17–36). Springer Berlin Heidelberg.

Tomasi, C., & Shi, J. (1994). Good features to track. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition CVPR-94* (pp. 593–600).

Trojanová, J., Křehnáč, K., & Brémond, F. (2016). Data-Driven Motion Pattern Segmentation in a Crowded Environments. In G. Hua & H. Jégou (Eds.), *Florida dental journal* (Vol. 9913, pp. 760–774). Cham: Springer International Publishing.

Wang, B., Ye, M., Li, X., & Zhao, F. (2011). Abnormal crowd behavior detection using size-adapted spatio-temporal features. *International Journal of Control, Automation and Systems*, *9*(5), 905–912.

Wang, C., Xu, Z., Yi, Z., & Yuncai, L. (2013). Analyzing motion patterns in crowded scenes via automatic tracklets clustering. *China Communications*, *10*(4), 144–154.

Wang, C., Zhao, X., Zou, Y., & Liu, Y. (2012). Unsupervised motion pattern mining for crowded scenes analysis. *KSII Transactions on Internet and Information Systems*, *6*(12), 3315–3337.

Wang, Q., Chen, M., & Li, X. (2017). Quantifying and Detecting Collective Motion by Manifold Learning. In *Proceedings of the 31th Conference on Artificial Intelligence (AAAI 2017)*, 4292–4298.

Wang, X., & Loy, C. (2017). Deep Learning for Scene-Independent Crowd Analysis. In *Group and Crowd Behavior for Computer Vision* (1st ed., pp. 209–252). Elsevier.

Wang, X., Ma, X., & Grimson, W. E. (2009). Unsupervised activity perception in crowded and complicated scenes using hierarchical bayesian models. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *31*(3), 539–555.

Wang, X., Yang, X., He, X., Teng, Q., & Gao, M. (2014). A high accuracy flow segmentation method in crowded scenes based on streakline. *Optik - International Journal for Light and Electron Optics*, *125*(3), 924–929.

White, S., & Smyth, P. (2005). A Spectral Clustering Approach To Finding Communities in Graphs. In *Proceedings of the 2005 SIAM International Conference on Data Mining* (pp. 274–285).

Wu, S., Moore, B. E., & Shah, M. (2010). Chaotic invariants of lagrangian particle trajectories for anomaly detection in crowded scenes. In *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2054–2060.

Wu, S., Su, H., Yang, H., Zheng, S., Fan, Y., & Zhou, Q. (2017). Bilinear dynamics for crowd video analysis. *Journal of Visual Communication and Image Representation*, *48*, 461–470.

Wu, S., & Wong, H. S. (2012). Crowd motion partitioning in a scattered motion field. *IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics*, *42*(5), 1443–1454.

Wu, S., Wong, H., & Yu, Z. (2014). A Bayesian Model for Crowd Escape Behavior Detection. *IEEE Transactions on Circuits and Systems for Video Technology*, *24*(1), 85–98.

Wu, S., Yang, H., Zheng, S., Su, H., Fan, Y., & Yang, M.-H. (2017). Crowd Behavior Analysis via Curl and Divergence of Motion Trajectories. *International Journal of Computer Vision*, *123*(3), 499–519.

Wu, S., Yang, H., Zheng, S., Su, H., Zhou, Q., & Lu, X. (2017). Motion sketch based crowd video retrieval. Multimedia Tools and Applications, 76(19), 20167–20195.

Yang, Y., Liu, J., & Shah, M. (2009). Video Scene Understanding Using Multi-scale Analysis. In *Proceedings of the 2009 IEEE 12th International Conference on Computer Vision*, (Iccv), 1669–1676.

Yi, S., Wang, X., Lu, C., Jia, J., & Li, H. (2017). L_0 Regularized Stationary-Time Estimation for Crowd Analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *39*(5), 981–994.

Zaharescu, A., & Wildes, R. (2010). Anomalous Behaviour Detection Using Spatiotemporal Oriented Energies, Subset Inclusion Histogram Comparison and Event-Driven Processing. *LNCS* (Vol. 6311 LNCS, pp. 563–576).

Zhang, W., Wang, X., Zhao, D., & Tang, X. (2012). Graph Degree Linkage: Agglomerative Clustering on a Directed Graph. *LNCS* (Vol. 7572 LNCS, pp. 428–441).

Zhang, Y., Qin, L., Zhao, S., Ji, R., Lu, X., Yao, H., & Huang, Q. (2016). Crowd video retrieval via deep attribute-embedding graph ranking. In *Proceedings of the 2016 IEEE International Conference on Multimedia and Expo (ICME)* (pp. 1–6).

Zhou, B., Tang, X., & Wang, X. (2012). Coherent filtering: Detecting coherent motions from crowd clutters. *LNCS*, *7573 LNCS*(PART 2), 857–871.

Zhou, B., Tang, X., Zhang, H., & Wang, X. (2014). Measuring crowd collectiveness. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *36*(8), 1586–1599.

Zhou, B., Wang, X., & Tang, X. (2011). Random field topic model for semantic region analysis in crowded scenes from tracklets. *Cvpr 2011*, 3441–3448.

Zhou, B., Wang, X., & Tang, X. (2012). Understanding collective crowd behaviors: Learning a Mixture model of Dynamic pedestrian-Agents. In *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2871–2878.

Ziliani, F., & Moscheni, F. (2007). *Kalman Filtering Motion Prediction for Recursive Spatio-Temporal Segmentation and Object Tracking*. *Workshop on Image Analysis, infoscience.epfl.ch*.

BIODATA OF STUDENT

Wong Pei Voon received the bachelor degree in Faculty of Computer Science and Information Technology from University of Malaya in 2004. She received the Master in Faculty of Computer Science and Information Technology from University Putra Malaysia in 2012. She is currently a part time Ph.D. candidate student in the Faculty of Computer Science and Information Technology from University Putra Malaysia. She is currently working as a lecturer in the Faculty of Information and Communication Technology, Universiti Tunku Abdul Rahman. Her research interests include data mining, social networks, and crowd behaviour analysis.

LIST OF PUBLICATIONS

Journals

Wong, P. V., Mustapha, N., Affendey, L. S., & Khalid, F. (2018). Collective Interaction Filtering Approach for Detection of Group in Diverse Crowded Scenes. *KSII Transactions on Internet and Information Systems*, 1-18. (Accepted)

Wong, P. V., Mustapha, N., Affendey, L. S., & Khalid, F. (2018). A Crowd Video Retrieval Framework Using Generic Descriptors. *Journal of Computers*, 1-12. (Accepted)

Proceedings:

Wong, P. V., Mustapha, N., Affendey, L. S., & Khalid, F. (2016). A new clustering approach for group detection in scene-independent dense crowds. *In 2016 3rd International Conference on Computer and Information Sciences (ICCOINS)* (pp. 414–417). IEEE. (Published)

UNIVERSITI PUTRA MALAYSIA

STATUS CONFIRMATION FOR THESIS / PROJECT REPORT AND COPYRIGHT

ACADEMIC SESSION : ______________________________

TITLE OF THESIS / PROJECT REPORT :

COLLECTIVE INTERACTION FILTERING WITH GRAPH-BASED DESCRIPTORS FOR CROWD BEHAVIOUR ANALYSIS

NAME OF STUDENT: WONG PEI VOON

I acknowledge that the copyright and other intellectual property in the thesis/project report belonged to Universiti Putra Malaysia and I agree to allow this thesis/project report to be placed at the library under the following terms:

1. This thesis/project report is the property of Universiti Putra Malaysia.
2. The library of Universiti Putra Malaysia has the right to make copies for educational purposes only.
3. The library of Universiti Putra Malaysia is allowed to make copies of this thesis for academic exchange.

I declare that this thesis is classified as :

*Please tick (√)

☐	**CONFIDENTIAL**	(Contain confidential information under Official Secret Act 1972).
☐	**RESTRICTED**	(Contains restricted information as specified by the organization/institution where research was done).
☐	**OPEN ACCESS**	I agree that my thesis/project report to be published as hard copy or online open access.

This thesis is submitted for :

☐	**PATENT**	Embargo from______________ until ______________ (date) (date)

Approved by:

______________________ (Signature of Student) New IC No/ Passport No.:	______________________ (Signature of Chairman of Supervisory Committee) Name:
Date :	Date :

[Note : If the thesis is CONFIDENTIAL or RESTRICTED, please attach with the letter from the organization/institution with period and reasons for confidentially or restricted.]