

University of Southampton Research Repository ePrints Soton

Copyright © and Moral Rights for this thesis are retained by the author and/or other copyright owners. A copy can be downloaded for personal non-commercial research or study, without prior permission or charge. This thesis cannot be reproduced or quoted extensively from without first obtaining permission in writing from the copyright holder/s. The content must not be changed in any way or sold commercially in any format or medium without the formal permission of the copyright holders.

When referring to this work, full bibliographic details including the author, title, awarding institution and date of the thesis must be given e.g.

AUTHOR (year of submission) "Full thesis title", University of Southampton, name of the University School or Department, PhD Thesis, pagination

UNIVERSITY OF SOUTHAMPTON

Application of Invariant Moments for Crowd Analysis

by

Hidayah Rahmalan

A thesis submitted in partial fulfillment for the
degree of Master of Philosophy

in the
Faculty of Engineering, Science and Mathematics
School of Electronics and Computer Science

January 2010

UNIVERSITY OF SOUTHAMPTON

ABSTRACT

FACULTY OF ENGINEERING, SCIENCE AND MATHEMATICS
SCHOOL OF ELECTRONICS AND COMPUTER SCIENCE

Master of Philosophy

by [Hidayah Rahmalan](#)

The advancement in technology such as the use of CCTV has improved the effects of monitoring crowds. However, the drawback of using CCTV is that the observer might miss some information because monitoring crowds through CCTV system is very laborious and cannot be performed for all the cameras simultaneously. Hence, integrating the image processing techniques into the CCTV surveillance system could give numerous key advantages, and is in fact the only way to deploy effective and affordable intelligent video security systems. Meanwhile, in monitoring crowds, this approach may provide an automated crowd analysis which may also help to improve the prevention of incidents and accelerate action triggering. One of the image processing techniques which might be appropriate is moment invariants. The moments for an individual object have been used widely and successfully in lots of application such as pattern recognition, object identification or image reconstruction. However, until now, moments have not been widely used for a group of objects, such as crowds. A new method Translation Invariant Orthonormal Chebyshev Moments has been proposed. It has been used to estimate crowd density, and compared with two other methods, the Grey Level Dependency Matrix and Minkowski Fractal Dimension. The extracted features are classified into a range of density by using a Self Organizing Map. A comparison of the classification results is done to determine which method gives the best performance for measuring crowd density by vision. The Grey Level Dependency Matrix gives slightly better performance than the Translation Invariant Orthonormal Chebyshev Moments. However, the latter requires less computational resources.

Acknowledgements

First and foremost, I would like to thank Allah the Almighty for giving me the strength and guidance to accomplish my thesis. Without His guidance, I will certainly not be able to complete this task successfully.

To my generous supervisors, Prof. Mark S. Nixon and Dr John N. Carter, thank you for allowing me to learn from my own mistakes, and for encouraging me to explore. I am truly grateful for the support, guidance, patience, and encouragement throughout this study and the writing up of this thesis. Your faultless guidance and unwavering trust will remain a standard of comparison for me in my future academic endeavors.

To my family, in particular my husband Norlizam who believes in what I am doing and for keeping everything together at home for all of us. My greatest appreciation goes to you. To my dear son Azzam and my lovely daughter Ilmuna, I thank you all very much for your patience, your cheeky manner, your many hugs and kisses and your loving energy that I needed most. To my parents and siblings, I especially thank you all for your many prayers during this endeavour. To all of you, without your love, faith and confidence in me, the process of completing this thesis would have been much more difficult.

It also gives me great pleasure to acknowledge and thank those who have helped, guided and supported me towards the completion of this degree. Those are :

- My Malaysian Southampton friends (Sheems, Syahida & Junaidi, Sis. Mi & Bro. Taib, Sis. Maya & Bro. Esa, Sis. Wana (Bro. Rahman / Bro. Shariman), Sis. Sham, Nanie & Azlan, Sis. Mahaza & Bro. Aiman, Sis. Umi & Bro. Ramizi, and to my dear sisters Adi & friends.)
- My ISIS friends (John Wynn, Tracey, Amanda, Ahmad, Imad, Houji, Anne, Nag, Shodan, Cem & Melica, Bassam, Dan, Stuart and Lee.)
- To all my colleagues in Utem especially to Prof Shahrin, Sis. Aniza, Sis. Chom, Dr Samad Shib., Lyana, all FTMK staffs and those who were involved with my research on crowd's data.

- To Azizah Saban and the Public Service Department of Malaysia for the moral and financial support.

Finally, my special thanks goes to thank En. Azman Abu, Dr Fazly Salleh Abas, Dr Mariana Yusoff and Dr. Salbiah, who had involved specifically or generally in fulfilling this thesis. May God the Almighty bless you all.

Contents

Abstract	i
Acknowledgements	ii
1 Introduction	1
1.1 Motivation : Why Crowds?	1
1.2 Aim of this study	2
1.3 Outline and Contribution	3
1.3.1 Outline of Report	3
1.3.2 Contributions	4
2 Crowds and Their Analysis	6
2.1 What is Crowd?	6
2.2 Why Crowd Analysis?	7
2.3 Crowd Analysis Application	10
2.3.1 Crowd Density	10
2.3.2 Crowd Movement	13
2.3.3 Crowd Behaviour	16
2.4 Summary	18
3 Moments: History and Contributions	19
3.1 Motivation: Why Moments?	19
3.2 Application of Moments	21
3.3 Analysis On Static Image	24
3.3.1 The Chebyshev Moments	25
3.3.2 Translation Invariant Orthonormal Chebyshev Moments	28
3.4 Summary	30
4 Experimental Data	31
4.1 Introduction	31
4.2 Data Organization	33
4.3 Boundary Binary Image	34
4.4 Summary	35
5 Measuring the Density of Crowds	37
5.1 Introduction	37

5.2	Overview of Experimental Methodology	38
5.2.1	Translation Invariant Orthonormal Chebyshev Moments	40
5.2.2	Grey Level Dependency Matrix	41
5.2.3	Minkowski Fractal Dimension	44
5.2.4	Comparison of Results	47
5.3	Further Analysis on Crowd Density Experiment	48
5.4	Summary	51
6	Conclusion and Future Work	53
6.1	Conclusions	53
6.2	Future Work	55
A	Sample Images For Measuring Crowd Density	57
	Bibliography	62

Chapter 1

Introduction

1.1 Motivation : Why Crowds?

The continuing development of surveillance technology has stimulated interest in analysing motion of people. This interest extends to the agglomeration of people: a crowd. This thesis concerns the analysis of crowds by computer vision.

A famous psychologist, Le Bon [35], concludes that a crowd is always intellectually inferior to an isolated individual, but, from the point of view of feelings and the acts these feelings provoke, a crowd may, according to circumstances, be worse than an individual person. All of these may depend on the nature of the environments to which the crowd is exposing to.

Consequently, the behaviour and pattern of crowds make a complex phenomenon. As described in [48]:

“ The aggregated motion [of a crowd] is both beautiful and complex to contemplate: beautiful due to the synchronisation, homogeneity and unity described in this type of motion, and complex because there are many parameters to be handled in order to provide these characteristics.”

Crowd analysis is also a multidisciplinary field. For example, researchers [19, 35, 55, 56, 60] are still studying the psychology of crowds. This group psychology may cause unpredictable situations in the crowd. Sociology is another field with deep interests in crowds as it concerns the patterns of social relationships and behaviour in the crowds. To ensure the crowds' safety, organizations involved with crowd management may be

concerned about handling and controlling crowds at events [22]. Meanwhile, those involved with designing buildings usually study crowds to provide a comfortable and safe environment [4, 22].

Initial studies [4, 6, 17, 49] show that monitoring crowds is a part of surveillance application. Furthermore, the advancement in technology has improved the effects of monitoring crowd, for example the use of CCTV. In addition, the use of CCTV has become an essential requirement in order to support a safe and secure operation continuously, especially in transport network or public services areas [6].

However, there are a few drawbacks of using CCTV. According to Boghossian and Black [6], the manual monitoring of live CCTV images and footage especially presents limitations linked to high maintenance cost, poor reliability of short attention span and lack of adequate training and experience of human operators. Furthermore, the observer might miss some information because monitoring crowds through CCTV system is very laborious[8].

One way to solve the problem above is by having the computer vision applications integrated into the CCTV surveillance system. This approach could give numerous key advantages, and is in fact the only way to deploy effective and affordable intelligent video security systems. In monitoring crowds, this approach may provide an automated crowd analysis which may also help to improve the prevention of unwanted incidents and accelerate action triggering. This was also supported by Boghossian and Black [6];

“...although current sensor technologies might offer direct solutions to address some of these issues, it is more attractive to address these issues at the image analysis stage of a vision based detection system. This is mainly in order to reduce possible overheads associated with replacing existing sensors in order to allow the current CCTV networks to benefit from computer vision solution. ”

1.2 Aim of this study

Being able to watch the CCTV system 24 hours for 7 days in order to prevent any crime or unwanted incident from occurring, especially among the crowds, is a difficult and challenging job to do. This problem has led the integration of computer vision techniques into the technology of CCTV system for a much better improvement in efficiency. Thus, the use of computer vision technique for surveillance application, especially in monitoring crowds has also increased.

Initially, this study began with a review of research works related to subject areas that has relevance to crowd analysis. To identify significant issues of interest, a review was made on multidisciplinary aspects of crowd, including views from computer vision. In general, crowd analysis can be observed through three important tasks such as crowd density, crowd movement and crowd behaviour.

In crowd density, the techniques of computer vision used were such as background subtraction [17], edge detection [17], texture [39], the Haar Wavelet Transform (HWT)[36], pixel counting [37] and the Kanade-Lucas-Tomasi feature (KLT) tracker[70]. Meanwhile, the techniques of computer vision applied in measuring the movement of crowd are such as block matching[8], Gabor filter [8], the Horn and Schunk algorithm[8], Kalman filter Kang et al. [31], and optical flow [3, 7, 8]. Since understanding the behaviour of crowds is a difficult task, researchers normally relate the behaviour of crowd with their movement. Examples can be seen in [7, 28, 59, 67].

According to Subbarao [62], human effectively perceive the shape and motion of unfamiliar objects as they manage to observe the changes of the objects' image. This shows that the changes of crowds can be measured using shape-based or motion-based techniques. However, Broggi et al. [10] stated that by applying the shape-based techniques, both moving and stationary pedestrians can be recognized.

In addition, if the objects' shape is represented as binary images, then the area of the objects can be easily calculated and used to describe the crowd density. Meanwhile, when there are changes between the shapes of objects, this can be used in measuring the movement of crowds. Thus, the motivation behind this study is to investigate a shape-based technique as one of the computer vision techniques to be applied in crowd analysis. More specifically, for this study, the monitoring crowd task was focused for estimating crowd density.

A series of images were used for the estimation of density. However, since there were no standard images of crowds available, a sequence of images was recorded specially.

1.3 Outline and Contribution

1.3.1 Outline of Report

This thesis is organized as follows:

- Chapter 2 presents the literature review of research on crowd analysis. It contains definitions of crowd terminology and presents views on crowds from studies on

computer vision and non computer vision. In the crowd analysis application, three important tasks: estimating crowd density; tracking a crowd's movement and understanding a crowd's behaviour were considered.

- Chapter 3 describes the current state of art and contribution related to moments which have been divided into analysis on static images. The contribution related to moments is called Translation Invariant Orthonormal Chebyshev Moments. It was modified from the Orthonormal Chebyshev Moments in order to provide moments that are invariants under translation.
- Chapter 4 describes the experimental data used in this study. Static images were used for crowd density. Samples of image for this experiment was also included.
- Chapter 5 describes the methodology and results of estimating crowd density. In this chapter, three different techniques were used and compared in order to obtain the best method to estimate crowd density. The three different techniques were the new Translation Invariant Orthonormal Chebyshev Moments, the Grey Level Dependency Matrix, and the Minkowski Fractal Dimension. Each image was labelled manually into one of 5 classes, from very low to very high range of density.
- Chapter 6 highlights the conclusions of this thesis and discusses directions for future research.

1.3.2 Contributions

The contributions associated with this work are

1. For many years, moments have been successfully used for a single object in lots of vision applications. Thus, the contribution of this research was to apply moments for groups of objects in which crowds were chosen as the problem.
2. Along with this, a taxonomy of crowds was provided.
3. A new algorithm Translation Invariant Orthonormal Chebyshev Moments was presented.
4. The comparison of TIOCM with MFD and GLDM has been carried out.
5. Papers presented on this research were:
 - Rahmalan, H., Nixon, M. S. and Carter, J. N. (2006) On Crowd Density Estimation for Surveillance. In Proceedings of International Conference on Crime Detection and Prevention, London UK, 13-14 June 2006. Organized by The Institution of Engineering and Technology (IET). (10 citation)

-
- Rahmalan, H., (2007) Using Invariant Moments for Crowds. The Inaugural STUDENT PAPERS MEETING, organized by BMVA, on 28th March 2007 at British Computer Society, London.
 - Rahmalan, H., (2008) Using Invariant Moments for Crowd's Density. The FTMK Colloquium, organized by FTMK, UTeM, on 10th September 2008 at Universiti Teknikal Malaysia Melaka, Malaysia.

Chapter 2

Crowds and Their Analysis

This chapter presents a taxonomy of crowds and a review related to crowd analysis. First, in section 2.1 some definitions of crowd terminology are considered, followed by a review of relevant studies with much interest in crowds in Section 2.2. Relevant studies based on a variety of disciplines which deal with crowds of people, include psychology, sociology, safety and military. Finally, a review on crowd application related to computer vision is presented in Section 2.3.

2.1 What is Crowd?

This section presents some definitions on crowds and terminology that has been used in crowd analysis. Gustave Le Bon, a famous French psychologist who wrote the book called ‘ The Crowd: A Study of the Popular Mind ’, describes a crowd as a gathering of individuals of whatever nationality, profession, or gender, and whatever be the chances that have brought them together [35]. Meanwhile, another definition on crowds by Macdonald [38], defines crowds as a temporary gathering of people who share a common focus of attention and who influence one another.

In recent years, video surveillance can focus on crowd monitoring. Furthermore, the use of CCTV allows us to monitor human behaviour, where the activities of persons or groups of people are observed without knowing their interaction. Unfortunately, as significant events are infrequent, human observers analysing crowd behaviour may lose concentration or might not even notice activity at all, or report it too late for an effective action.

Hence, there is much interest in automatic crowd analysis. Crowd analysis is often concerned with group behaviour such as preferential motion direction and magnitude

[17]. Davies et al. [17] also state that engineers should take this opportunity to suggest solutions to crowd monitoring and control, based on technological developments in image processing or image understanding. In this thesis, the primary concern is the measure of crowd density, which may be defined as :

- Overcrowding can be defined as a situation in which an excessively large number of people are gathered within a defined area and this could lead to injuries and, in extreme cases, death [4].
- The safety limit for crowd density is defined as $4\text{people}/m^2$ if they are moving and $4.7\text{people}/m^2$ if they are stationary.
- Hughes [30] defines that the density of a crowd is generally measured by the number of pedestrians found per unit area.

However, there are other areas of study in the literature. These include :

Movement and Dynamic is the study of how people move and their interaction with other people closed to them.

Management and Control is defined as planning for the movement and gathering of people, in order to implement the avoidance of critical crowd densities and the triggering of rapid group movement Fruin [22].

In addition, people have considered how communication, mobility, building design and the level of service have being studied.

2.2 Why Crowd Analysis?

Crowd analysis is a subject that can be viewed from different disciplines. In this section, aspects of crowds are discussed from different perspectives such as from psychology, sociology, safety and military. The first discipline that will be presented is crowds from the psychological perspective. Psychology is a science in which behavioural and non-behavioural evidences are used to understand the internal process leading people and members of other species to behave as they do [19]. Thus, crowd psychology is meant to understand the minds and behaviour of the people who cluster together.

As crowds usually involve large groups of people, ordinary people can typically gain direct power by acting collectively. Historically, this situation can cause dramatic and sudden social change in a manner that may appear violent, especially when people

become mindless. This idea was introduced by Le Bon [35], in which his work is still widely cited today and has been called as the most influential psychology text of all time. He had argued that when a person becomes anonymous within the mass, he will tend to lose his individual identity. This may cause that person to forget his normal values and standards including his ability to think. Thus, being one of the crowd members, when there is a suggestion which are normally may consider and reject, it is no longer resist and may become effectively mindless. Some may then act heroically, or even worse, some may act like primitive beings, with barbaric and destructive behaviours.

Similar to Le Bon [35], Global Security [24] state that the emotional contagion provides the crowd psychological “ unity ” and may give a momentum in which a crowd will turn to mob action. A mob is a crowd that often acts unreasonably with potential violence, especially when a crowd follows its leaders into unlawful and disruptive acts. As emotional contagion prevails, the self-discipline of crowd members is low and normal controls give way to raw emotions. With no fear, the member of crowds may release their personal prejudices and unsatisfied desires which they wanted to do, but dared not try alone.

Unlike Le Bon [35], Reicher [55] strongly disagrees that crowds can be seen as anonymous, with their actions inherently destructive and random, and their reasons unfathomable. Reicher [55] believes that the crowd’s action can be patterned in such a way as to reflect their existing cultures and societies. His observation on crowds is based on a sociological perspective because sociology is a scientific study of human social life, groups, and societies [23].

In order to understand how crowd events and crowd conflict developed over time, [55, 56] aimed to extend the social identity approach to crowd behaviour. This study concluded that, while the social identity model is of use in understanding the crowd phenomena, it is necessary to recognize how social categories are constructed and reconstructed in the dynamics of inter-group interaction. The social identity can be thought of as a function of what group the crowd members are involved in. [55, 56] gave an example as such that the values and standards or behaviour of crowds involving university students are very different from a crowd of soccer supporters. Consequently, understanding the social identity in the crowds is essential as the behaviour of different crowds will vary.

Another factor that can influence the crowd’s behaviour is communication [22]. Due to a lack of communication, people in the rear of a crowd may press forward until those in front experience severe distress become immobile and are under great pressure. The tragedy of Heysel Stadium which happened in 1985 as reported by Chisari [15] occurred due to lack of this kind of communication.

However, there were also a lack of communication among the authorities as is desired in the following quotation.

“ The Belgian authorities had been completely unable to manage the emergency and, even more seriously, had underestimated the security required for such a delicate and potentially incendiary event. Thus, after the initial rage at the behaviour of English hooligans, Italian and international public opinion began to point fingers at the Belgian police and Gendarmerie and at the Minister of Interior, Mr Charles-Ferdinand Nothomb. The police were accused of making mistakes in the deployment of forces in the crucial hours before the game, with a subsequent lack of communication among the units, and of poor professional behaviour in the face of the emergency.”
Chisari [15]

However, in a situation where authorities have been unprofessional, combination factors of psychology and sociology are important to ensure that crowds are safe from terrible incidents. A recent example happened in Malaysia, as reported by BBC news [9]

“ Malaysian police have used tear gas and water cannon to disperse thousands of protesters who were marching in Kuala Lumpur to demand electoral reform. The event was organised by a group called Bersih which is made up of opposition parties and dozens of non-governmental organisations. The marchers were prevented from entering central Merdeka Square because police said they did not have a permit. Police estimated the crowd at between 10-30,000 people.”

Psychologically, from the example above, crowds can be in a peaceful situation when the crowd's members have positive mindset and remain patient even if provocation occurs. Meanwhile led by a good leader of high moral attitudes and self-discipline among crowd members, and good communication skills are some sociological factors that help crowds to remain safe. And when the police did not play a sufficient role to control the crowds, a number of brave people among the crowd members, in this example are the Badan Amal team, and should come forward and protect the crowd members from any harm or accident. This was reported by a Malaysiakini journalist [63] who wrote :

“ The unsung heroes of the day were undoubtedly the Badan Amal team from PAS. The maroon-clad army played a crucial role at the rally, not just directing crowds and traffic. These brave men were right in the thick

of things, protecting frontline protesters and pulling people to safety when clashes occurred. ”

Meanwhile, the measures used to effect control in a disturbance will affect the crowd’s behaviour. This is because each crowd is unique and may vary during the disturbance, as described in [24]:

“ Control force measures must be geared to each crowd’s size, temperament, cooperativeness, and degree of organization and uniformity. Measures should change as the crowd’s characteristics change. Even a change of one characteristic can drastically alter a crowd’s response to control force measures. Large crowds may be easy to control if they are organized, cooperative, and non-violent. Non-violent crowds are often easy to control with a very limited show of force. Small crowds can be hard to control if they are organized, uncooperative, and violent. ”

The examples above shows that had an effective automatic measure of crowd density, movement or behaviour than the incidents might have been detected earlier and those of wroted.

2.3 Crowd Analysis Application

This section presents some research related to crowds and their measurement.

2.3.1 Crowd Density

This section presents research on crowd density from the perspective of computer vision. In general, crowd density can be defined as the number of people per unit area. The importance of measuring crowd density is due to several reasons. In [4], one of the key aspects in developing and maintaining a crowd safety system is to identify areas where crowds build up. Areas where crowds are likely to build up should be identified prior to the event or operation of the venue. This is important as crowds usually exist in certain areas or at particular times of the day. Areas where people are likely to congregate need careful observation to ensure crowd safety. Therefore, estimating crowds’ density is a good solution for maintaining a safety environment for crowd.

Nicholson [49] states that, automatic estimation of crowd density is crucial for automatic monitoring of crowds since when a crowd exceeds a certain level, danger may occur for

several reasons. People's safety may be endangered when the area of crowds reaches an occupation level greater than the projected one [42]. Marana et al. [42] also state that in order to prevent accidents, mainly in routine monitoring, such as those carried out in airports, train and subway stations, automatic crowd density estimation is needed.

Furthermore, crowd density and crowd movement can be related. This is suggested by Fang et al. [20] who said :

“ The crowd flow pattern under emergency situations has been studied on the basis of dynamic movement principles. It has been demonstrated that the surrounding crowd density will influence the speed of an individual. The derivation of the movement equations of people with respect to the impacts at front, back and lateral directions has been given, and it shows that the impacts have substantial influence on the peoples' speed. The logarithmic relationship between the crowd density and the speed is in good agreement with the published field data. The study also demonstrates that the influence of the crowd movement velocity by the inter-person effect at the lateral direction is much lower than that at the frontback direction. ”

The above statements are a great motivation to estimate crowd density. Moreover, research work of crowd density related to computer vision techniques is presented in this section. Perhaps the earliest analysis on crowds using image processing approach was proposed by Velastin et al. [68]. Davies et al. [17] proposed a technique to estimate the crowd density based on two measurements extracted from the input image of the area under surveillance. The first measurement measures the number of foreground picture elements, computed by subtracting the input image from a reference image containing no people; while the second measurement measures the number of edge picture elements of the image computed by edge detection, followed by a thinning operation. They verified that there exists a linear relationship between the number of people present in the area under surveillance and the two measures, which were combined into an optimal estimate of crowd density through a linear Kalman filter [17].

Even though this technique was successful in estimating crowd density, it cannot however be applied successfully in areas with high density crowds. This is because the linear relationship does not hold when there is occlusion or when people overlap in the image.

As an alternative, Marana et al. [39] came with a suggestion to divide the crowds estimation into a range of classes, from very low to very high density. To achieve these results, texture information based on the Grey Level Dependency Matrix (GLDM) method were used on digitised images. Ideally, the technique identifies images of low density crowds

as coarse textures, while images of high density crowds are presented as fine textures. This method was also compared with other methods, such as Fourier spectrum and Minkowski fractals [39, 40, 42]. Among these 3 methods, the GLDM performs well as it can provide classification performance above 80%. However, through our observation, these are the only methods that provide crowd density estimation into a range of density classes.

Meanwhile, Lin et al. [36] proposed a technique that counts people in crowds based on their head-like contour. First, the Haar Wavelet Transform is used to extract the featured area of the head-like contour, and then a Support Vector Machine is used to classify these featured area as the contour of a head or not. After all possible locations of people are detected, the information sizes and positions of the detected frames of people are used to estimate the number of people in crowds with a perspective transformation [36]. Even though this technique works, unfortunately it is time consuming as it considers detailed information, due to the head-like contour and location of each person in the crowd.

Due to many occlusions, counting crowds is difficult, even with strong prior assumptions and no computational limitations, and it is impossible to count the crowd from a single view. But, it may be possible by using many sensors in a sensor network. Yang et al. [74] developed a system that counts people in a crowded scene using a network of simple image sensors. They proposed a geometric algorithm that computes bounds on the number and possible locations of people using silhouettes computed by each sensor through background subtraction. The system requires no initialization, runs in real-time and does not compute any feature correspondence across views. However, the computation cost increases linearly with the number of cameras.

Ma et al. [37] applies a pixel counting method that relies on the foreground segmentation result. However, the pixel counting based method is substantially affected by perspective distortion. They then investigated a geometric correction to account for the perspective distortion, where the geometric correction has the possibility to bring all the objects at different distances in a scene to the same scale. Normally the derived transform for geometric correction is valid for the ground plane but not the human bodies, an issue that has been neglected by previous work, according to Ma et al. [37]. Applying the geometric correction to human bodies leads to a linear relationship between the number of pixels and the number of persons regardless of their relative position in the scene, without more complex considerations of 3D information.

Later in 2006, Vincent and Serge [70] introduced a trajectory set clustering method to identify the number of moving objects in a scene. By using the Kanade-Lucas-Tomasi feature (KLT) tracker [70], it can populate the spatio-temporal volume with a large set

of feature trajectories efficiently. After applying an algorithm that can smoothen and extend the raw feature trajectories, the conditioned trajectories were then clustered into a candidate object using local rigidity constraints and a simple object model learned from a small set of training frames labelled only with the ground truth object count. The method was applied on three different datasets [70]; 1) USC dataset consisting zero to 12 persons of crowd; 2) Library dataset consisting 20 to 50 persons of crowd and 3) Red blood cell dataset with 50 to 100 blood cells moving at different speed. Meanwhile the average error in percentage for these datasets were 10, 6.3 and 22 for USC, library and cells datasets respectively. Unlike most other approaches, this method does not require background subtraction.

2.3.2 Crowd Movement

Another interesting application on monitoring crowds is observing the crowd's movement. Crowd movement not only has a logarithmic relationship with the crowd density [20], but also help the crowd management team to recognise any potential emergency situation [32]. Besides that, in transportation systems, movement of crowds may influence the design and operation of transportation terminals and timing of traffic signals [5]. [32] divided the movement of a crowd into a general classification based on a control,

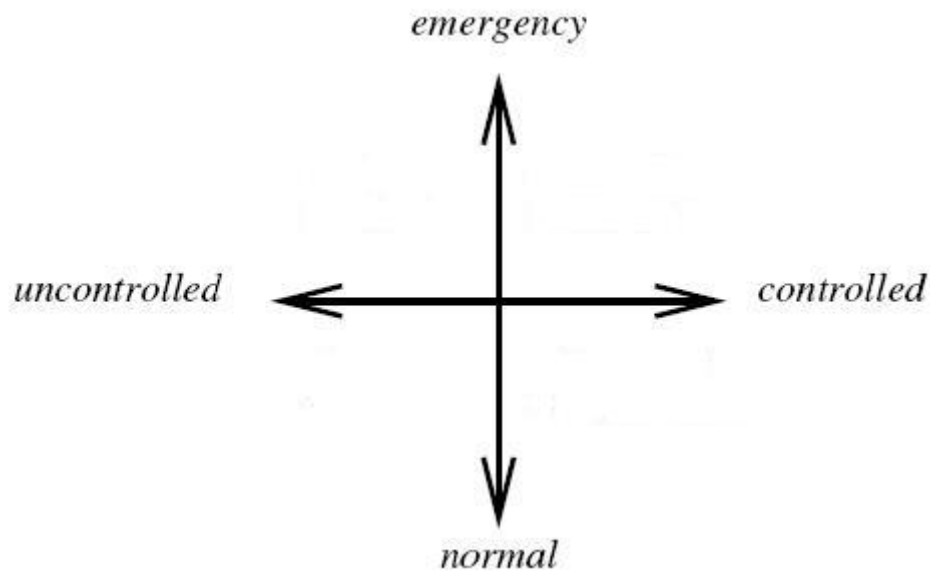


FIGURE 2.1: A general classification of crowd movement taken from [32]

uncontrolled, emergency or normal situation. The classification of a crowd's movement is displayed in Figure 2.1. In reality, the uncontrolled movement situation is considered

the real situation that needs to be observed, in order to prevent any accident or abnormal behaviour regarding the crowd's safety. This is where CCTV and automated crowd monitoring using image processing techniques are of great interest.

Meanwhile the controlled movement situations such as evacuation exercises and movement experiments were usually related to further research on movement. Usually, developing a simulation may provide a clear and safe way for the evacuation of people in complex environments. According to Musse et al. [48], there are also other reasons for developing simulation of a crowd's movement: 1) entertainment purposes such as developing animation film or game, 2) to populate collaborative and distributed virtual environments in order to increase the credibility of people's presence and 3) to analyse sociological and behavioural views, such as the relationship between different people, or the hierarchy existing inside a crowd.

To learn or investigate the movement of crowd, there are two types of model that can be used. These models are called the microscopic or the macroscopic approach [1, 2]. The microscopic approach involves a detailed design that focuses on each individual object and their characteristics such as individual interactions, direction, and speed. The movement of each individual object in this type of model is based on changes in their surrounding environment. Legion is an example of this model, developed by Still [61] which describes the dynamics of crowd. The model treats every entity as an individual and it can simulate how people read and react to their environment in a variety of conditions, which also allows the user to study a wide range of crowd dynamics in different geometries and highlights the interactions of the crowd with its environment. However, the drawback of microscopic models is that they are computationally expensive especially when the number of objects increases.

On the other hand, the macroscopic approach concerns group behaviour and it deals with a crowd as a whole. Macroscopic modelling is well suited to understanding the rules governing the overall behaviour of pedestrian flow for which individual differences are not that important. The detailed interactions are ignored and the model's characteristics are shifted towards continuum density flow and average speed [2]. Thus, this has been the advantage for macroscopic approach as they are computationally less demanding than the microscopic approach.

Considering a macroscopic modelling approach, [2, 13] are simulation examples of pedestrian flow based on numerical studies. Some examples of computer vision techniques in relation to crowd movement analysis from a macroscopic approach are written below.

Motion-based algorithms can be used to detect the movement of crowds, related to emergencies and to assist CCTV operators in ensuring crowd safety. There was an example written by Boghossian and Velastin [7]:

“ To improve the system performance and crowd safety, there were three detections of potentially dangerous events being observed: 1) detection of circular flow paths, which originate close to scene exits when large crowds attempt to evacuate the scene. Scene exits usually act like a bottleneck for large crowd flow, thus when different flows meet at the exits, some are pushed back into the scene creating circular flow paths; 2) detection of diverging flows happen when there are fights or fires which are manually detected by spotting the diverging crowd flow from the location; and finally 3) detection of obstacles, which were visualised as motion free regions surrounded by homogeneous flow.”

The movement types described above could be categorized as self-organized phenomena [28], where self-organized phenomenon can be described as a collective pattern of motions which was not externally planned, prescribed or organized but by the spatiotemporal patterns that emerge due to the non-linear interactions of pedestrians.

Bouchafa et al. [8] provide an automatic tool for the detection of abnormal individual or crowd motion in subway corridors. They developed a motion estimation method that takes into account two difficulties: real time constraints and non-rigid moving objects. By using three different techniques such as block matching, Gabor filter and the Horn and Schunk algorithm, they attempt to improve both the speed and performance of each algorithm. Unfortunately none of those algorithms work at the required speed. The optical flow technique was then modified to give satisfactory results in real time. Later, the obtained motion vectors were then filtered and classified into two groups (either having a correct direction or an opposite direction). These results were then used for the construction of motion trajectories in order to analyse the behaviour of the crowd.

Next, Kang et al. [31] proposed an approach that consists of modelling the multiple trajectories observed by the moving and stationary cameras in the same Kalman Filter framework. The detection of the moving objects from the moving camera is performed by defining an adaptive background model that takes into account the camera motion approximated by an affine transformation. It will address the tracking problem by separating the motion and appearance estimation of the moving objects using two probabilistic models where: 1) for the appearance model, multiple colour distribution components were proposed to ensure a more detail description of the object being tracked; and 2)

for motion model, a Kalman Filter was used to predict the position of the moving object. The tracking moving objects were then performed by the maximization of a joint probability model.

Andrade et al. [3] developed an event detector for monitoring emergency situations in crowds. This is done by learning patterns and comparing the normal crowd flow with two emergency scenarios: 1) the simulation of a blocked exit in the scene and 2) a person falling on the floor. Here, optical flow was used to obtain the motion information and these features are encoded with Hidden Markov Models to allow for the detection of emergency or abnormal events in the crowd.

Meanwhile, Courty and Corpetti [16] proposed a data-driven animation scheme that allows generating animation of crowds from a video of a real crowd. This animation model has been tested on a synthetic sequence, representing a continuous flow of human beings with an obstacle and also two real sequence images of crowds having strike and entering a stadium. Information on motion of crowds were presented as a time series of velocity fields estimated from the video and were used as an input to an animation model that advects people along this time varying flow. The results show that the motion information can infer the presence of the complex behaviour or scene semantics of crowd. In addition, some applications which analyse crowd's behaviour using computer vision techniques will be presented in the following section.

2.3.3 Crowd Behaviour

The most difficult task when monitoring a crowd is to understand its behaviour. This is difficult from the psychological or sociological perspectives, as understanding crowd's behaviour not only requires physical appearance but also added with moral value to describe the humans' behaviour which is intangible to computerise. Hence, it appears that the crowd's behaviour is very complex to be computerised by computer vision techniques. Yet, learning and monitoring crowd's behaviour are important since it is essential for the authorities such as police, to investigate any suspicious activity or crime scene within the crowds. Therefore, research work on crowd behaviour is still an ongoing process.

As a part of the study of crowd behaviour, research has also concentrated on the people's activity. Haritaoglu et al. [27] constructed a real-time system for detecting and tracking people and their body parts in video imagery. They classified humans and their activities by detecting features such as hands, feet and head, and then tracking and fitting them to a prior human model. The system performed a dynamic appearance model by combining the gray-scale textural appearance and shape information of person in a 2D dynamic

template, called a textural temporal template, to recognize people's action. This can be used to determine the correspondence between the people who were tracked before the interaction, and the people who emerge from the interaction. However, the primary purpose of the research was not for crowd monitoring.

On the other hand, investigations on the behaviour of pedestrian flow in crowds has been done by developing a lattice-gas model [76]. This was done for the purpose to study the dynamical phase transition and scaling behaviour of crowd outside a hall. Even though this may help to understand the pedestrians' behaviour, it is not practical for certain cases which involve highly crowded scenes, such as emergency evacuations.

Conversely, Vaswani et al. [67] proposed a model that learned the pattern of normal activities and detected abnormal events from very low resolution video. They used a particle filter to estimate a probability density function for the actual shape given the noisy observations up to the current time. To detect abnormality, [67] proposed comparing the distance of the estimated probability density function from the probability density function learnt earlier for a normal activity, using Kullback-Leibler distance. The Kullback-Leibler distance is also known as relative entropy [72] while a particle filter is a recursive algorithm that approximates, by Monte Carlo sampling, the optimal posterior distribution of the state at any time given the past observations.

In 2004, Siebel and Maybank [59] developed a system named as ADVISOR. This system combines four co-operating detection and tracking modules: Motion Detector, Region Tracker, Head Detector and Active Shape Tracker to track people in camera images. The objective of the ADVISOR system is to detect and classify anomalous events by learning the behaviours of individuals and crowds for metro stations. This suggests that crowd movement can be used to describe the crowd behaviour.

In addition, Helbing et al. [28] developed a simulation to improve the self-organized phenomenon in order to describe the crowd behaviour by observing the patterns of crowd movement by suitable specification of the boundary conditions. This sort of obstacle or boundary, if located in a suitable place, may stabilize the flow patterns of a crowd movement and make them move much smoother, like a fluid.

Although human's behaviour in crowds are usually described based on microscopic approach as such observing each individual's behaviour or action, the self-organized phenomenon has shown that macroscopic approach can also be another way of observing the crowds' behaviour.

2.4 Summary

A taxonomy and studies relevant to crowds have been presented. The crowd analysis as described was viewed from fields other than computer vision. Analysis of crowd has shown that studies from psychology or sociology can contribute to engineering study. The three most important tasks on monitoring crowds appear to be: estimating crowd density; tracking the crowd's movement; and observing crowd's behaviour.

A selection of computer vision approaches has been used within crowd analysis, such as edge detection [17], texture description [42], and optical flow [3]. However, this has not been done using the statistical moments, which will be considered later. In the updated survey on crowd analysis by Zhan et al. [77], the contribution of this study (TIOCM) was the only research work that has been using moments as a technique of computer vision in crowd analysis application so far.

In addition, even though observing crowd behaviour was the most interesting task in crowd monitoring, let us not forget that it cannot be done without the information based on crowd density and crowd movement. Thus, this study using moments on crowd will start its investigation on crowd density and crowd movement. In addition, the following chapter will then highlight experimental data for both tasks.

Chapter 3

Moments: History and Contributions

The previous chapter described that computer vision technique is a solution to improve the efficiency of using the CCTV technology. Chapter 2 has reviewed several techniques of computer vision applied in crowd analysis. However, moment invariants has been lack in used for groups of objects such as crowds. Thus, in this study, it became an opportunity to investigate crowds using moment invariants as one of the computer vision technique.

In this chapter, more information about moments is presented. This includes a general study using moments and also a contribution related to moments. Section 3.1 gives a brief description of the motivation on using moments. The following section describes the application of moments. Later, analysis on static images using moments will be described in sections 3.3.

3.1 Motivation: Why Moments?

In vision applications, an object can be represented as a collection of pixels of an image. For the purpose of recognition, the object can be described by a set of numbers which represent the properties of the object. These properties are called the object's descriptors. There are two common forms of descriptors which are i) region and ii) boundary [51]. Among these descriptors, moments are categorized into region descriptors.

The advantages of moments are that they are compact in description and are capable of selecting different levels of detail [53]. For example, when applied to binary image, the zero order of moments can be interpreted as the area, the second order of moments

can be interpreted as the variance, the third order of moments can be interpreted as skewness while the fourth order of moments can be interpreted as kurtosis [46]. As a result, this has made moments a global shape descriptor [51] and a powerful statistical tool for pattern recognition [78]. Thus, by using moments, different features can be used to describe the object extracted from an image.

Recognizing an object using moments, especially with added capability concerning independence of position, size or orientation, has been a goal for this research area until now [29, 46, 53, 58]. In fact, the earliest pattern recognition with the use of moments in computer vision applications dates back to the pioneering work of Hu [29]. By using geometric moments, he generated a set of moments which was invariant under translation, scale and rotation for automatic character recognition. Later works have improved and generalised Hu's invariant moments.

According to Shutler [58], moments can be divided into; 1) non-orthogonal moments, and 2) orthogonal moments. An example of non-orthogonal moment is Geometric moments which can be defined as the projection of a function $f(x, y)$ onto a monomial $x^p y^q$. However, the higher order of geometric moments are more sensitive to image noise and can lead to mismatch in pattern recognition algorithms [46].

As a solution to the problem above, Teague [64] introduced Legendre and Zernike polynomials as a kernel function to the orthogonal moments. These orthogonal moments can be used to reduce the number of redundant features in a moments set, so that the moments correspond to independent characteristics of the image.

Regardless of their advantages, both types of moments (orthogonal and non-orthogonal) have been applied to many application such as image analysis [65], texture segmentation [66], multispectral texture [71], pattern recognition [58, 81], image watermarking [79] and image reconstruction [44, 45, 52, 53].

According to Xu and Li [73], there are mainly three research directions or extensions based on the regular moment invariants ;

1. Observing moment invariants in arbitrary dimensions, from 2D, 3D to nD. Observation using 2-D moment invariants has been successfully applied in vision application , but extention to high dimensions are still few.
2. Using moment invariants under different transformations as such affine, translation, uniform scaling and rotation.
3. Applying moment invariants on different manifolds or domains. For example, in 3-D space, there are mainly three kinds of manifolds.

Based on this review, in most vision applications, moments have been successfully used to recognise a single object. However, their application to a group of objects in a scene, namely crowds, has yet to be performed. By taking these facts into consideration, the aim of the present study is, therefore, to accrue the advantage of a moment description when applied to analysing a group of moving objects, namely crowds.

3.2 Application of Moments

This section presents an introduction to computational shape analysis using moment invariants. This explanation is intended to give an overview on using moments for recognition applications. Basically, the first process starts with taking an image $f(x, y)$ of the target object. This will be then followed by applying feature extraction. There are many vision techniques that can be used to extract the features and one of them is moment invariants. After the feature extraction process, it shall be followed by the classification process in order to recognize the target object. In this explanation, the evaluation of moments is applied to a single object silhouette.

Among all types of moments, geometric moments are known as the simplest moments[53]. The two-dimensional geometric moment of order $(p + q)$ are denoted by m_{pq} , and can be expressed as

$$m_{pq} = \int \int_I x^p y^q f(x, y) dx dy \quad (3.1)$$

where $p, q = 0, 1, 2 \dots n$ and I is the region of the pixel space in which the image intensity function $f(x, y)$ is defined. Note that the monomial product $x^p y^q$ is the basis function for this definition.

Due to the discrete nature of a computer, the moments from an image cannot be directly calculated from equation 3.1. As a result, equation 3.2 can be used as an approximation to equation 3.1.

$$m_{pq} = \sum \sum_I x^p y^q f(x, y) \quad (3.2)$$

It appears that the most commonly used region attributes are calculated from the three moments of lowest order: m_{00}, m_{10}, m_{01} . The zeroth order moment m_{00} written below represents the total intensity of the image, $f(x, y)$ [46]. When the image represents the sum of binary pixels interior to the region, it will then give a measure of the area.

$$m_{00} = \sum \sum_I f(x, y) \quad (3.3)$$

The first order moments m_{10} and m_{01} are given by

$$m_{10} = \sum \sum_I x f(x, y) \quad , \quad m_{01} = \sum \sum_I y f(x, y) \quad (3.4)$$

These first order moments m_{10} and m_{01} provide the moments intensity according to the x and y axis of the image respectively. It is also useful to measure the centroid of the object. The centroid or the Centre Of Mass (\bar{x}, \bar{y}) , can be calculated from the ratio of the first order moments to the zeroth order moments as written below:

$$\bar{x} = \frac{m_{10}}{m_{00}} \quad , \quad \bar{y} = \frac{m_{01}}{m_{00}} \quad (3.5)$$

The higher order of geometric moments is not significant for describing the moving object. This is due to the formulation of moments (as written in equation 3.2) as it depends on the value of x or y axis. For example, if the original object moves towards higher value of x axis, then the geometric moments with respect to the x axis of the particular object will increase. As a result, the geometric moments produce large numbers and this will produce difficulties in detecting the moving object.

Thus, as an alternative, geometric moments can be developed to be central moments which have the property of invariance under translation. The definition of central moments μ_{pq} is subtracting each pixel of the geometric moments equation with the centroid as written below:

$$\mu_{pq} = \sum \sum_I (x - \bar{x})^p (y - \bar{y})^q f(x, y) \quad (3.6)$$

The central moments provide significant measurement for describing moving objects as the moments will remain the same even when the objects have changed their position (moved). This happens as the image centroid (\bar{x}, \bar{y}) moves with the image under translation, and the central moments are defined with respect to this point as the origin [46]. In addition, central moments property that are usually used are the second, third and fourth order moments. The second order of central moments (μ_{20}, μ_{02}) for an image intensity distribution measures the variance about the origin, while the covariance can be measured by giving μ_{11} [46].

The third (μ_{30}, μ_{03}) and fourth (μ_{40}, μ_{04}) order of central moments can be thought as skewness and kurtosis respectively. Skewness characterizes the degree of asymmetry of a distribution around its mean while kurtosis is a degree of peakedness of a distribution. Some researcher use skewness as an approach to locate the features of an object's shape rather than relying on finding the border of the particular object's shape [51]. Another example by [25], used statistical formula such as skewness and kurtosis to differentiate between art images and natural real scenery. He found that the intensity distributions of art images show lower skewness and kurtosis compared to natural scene, as there were significant differences in the mean amplitude spectrum slopes.

Meanwhile, in applying image reconstruction, the higher order of orthogonal moments becomes really useful. A good example of image reconstruction can be seen in [53] and [47]. However, in this section, a simple example on the use of the zeroth and second order moments were presented. Figure 3.1 displays an example of an object's silhouette with three different positions where there were an original image, a rotated image and a translated image.


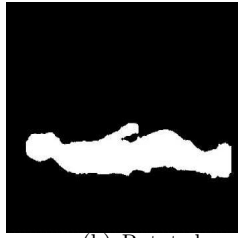
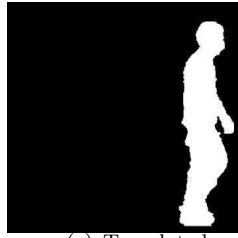
image	 (a) Original	 (b) Rotated	 (c) Translated
μ_{00}	19472	19472	19472
\bar{x}	138.81	214.13	338.81
\bar{y}	214.13	262.19	214.13
μ_{20}	7.22E+06	1.74E+08	7.22E+06
μ_{02}	1.74E+08	7.22E+06	1.74E+08

FIGURE 3.1: Image of a Silhouette

Since the same object was used in Figure 3.1, the area as represented by μ_{00} remained the same in each image. Note that the result taken from μ_{00} and m_{00} are similar. The \bar{y} in the original image is equal to the \bar{x} of the rotated image as the silhouette subject was rotated 90° from the original image. Meanwhile the \bar{y} in the translated image remained constant such as in the original image since there were no changes in y co-ordinate of the subject. However, the \bar{x} of the translated image was increased since the subject has moved 200 pixels to the right. Interestingly, μ_{20} and μ_{02} of the original and translated image were equal, showing that the same subject could be recognized even if they were moving and the central moments, numerically were invariant under translation.

In handling a set of measurement of data, some vision applications will require a classification process to assign these measurements into certain class labels. If there were a

number of objects to be recognized and classified, then the measurement of these objects taken from the images will need to be delivered to a classifier. The classifier will organize the feature space into regions corresponding the class labels [12]. Some examples of classifier are SOM, K-means and K-nearest neighbour [12, 33, 51].

Before classification, the moment features were computed from the images, irrespective of the position and orientation of the image in its plane. Then these features will be divided into two: 1) Training data (the prelabel data) which is a description of features of the referred class object sample and 2) Test data (the unknown label data) which is the description of features that will be used for classification. The classification shall be done according to the element of the database that best matches the classes in the training sample data.

If the training data were classified based on their class labels, then the classification process is termed as supervised learning classification. The following step is to match the test data with the data that has been supervised. On the contrary, if the training data were classified without any class labels, then the classification process is called unsupervised learning classification. The process of unsupervised learning classification will maximize the similarity of features within the objects in the same class and minimize the similarity of features between objects in different classes [12]. Each group of objects is called a cluster. The object that belongs to a particular cluster will be labelled according to that cluster.

3.3 Analysis On Static Image

One of the interesting task in crowd analysis is estimating crowd density. Instead of providing the exact number of people in the crowd, [39, 41, 75] provide the crowd estimation as a range of density using texture method. In general, if the image looks like a fine texture, then the output can be interpreted as high density, but if the image looks like a coarse texture the output can be of low density.

This gives an idea that moments may also be able to provide the estimation of the crowd density into the same range of class density above. For example, it is predicted that if the zero order moments produce high value, then it will be interpreted as high density, while small value can represent low density.

Rather than using the non-orthogonal moments, it is advisable to use orthogonal moments. This is because, according to Shutler [58], the monomials basis set of the non-orthogonal moments such as geometric moments (x^p, y^q) increases rapidly in range as

the order increases. This may result in important descriptive information being contained within small differences between moments, in which it will lead to the need for high computational precision.

Meanwhile, the moments that are produced using orthogonal basis sets, will have the advantage of needing lower precision to represent differences to the same accuracy as the monomials. Due to this advantage, the orthogonality condition usually simplifies the reconstruction of the original function from the generated moments [58].

In the year 2000, a new orthogonal moment named as Chebyshev moments (also known as Tchebichef moments) was established. The Chebyshev moments were introduced by [43] to improve the results of image reconstruction. The Chebyshev moments were discrete moments and are still interests many researchers[54, 79, 81]. Besides, the discrete orthogonal Chebyshev moments have some advantages when compared to the continuous conventional orthogonal moments such as Legendre and Zernike moments [43]. Thus the Chebyshev moments were chosen in this study as it was the most recent orthogonal moments used in vision application. Section 3.3.1 presents the original Chebyshev moments that were referred to in this study. Meanwhile section 3.3.2 presents the new contribution of this study with a modification from section 3.3.1.

3.3.1 The Chebyshev Moments

Based on [43] observations, the discrete orthogonal Chebyshev moments are much better than the other conventional orthogonal moments such as Legendre and Zernike moments due to; 1) the discrete orthogonal moments provides a better result compared to conventional moments as it does not require a numerical approximation. 2) the discrete orthogonal moments satisfies the orthogonality property and can be defined in the image coordinate space. 3) using discrete orthogonal moments, the reconstruction images have better quality.

Later, [44] found that the discrete orthonormal Chebyshev moments appear to give a much better performance compared to the discrete orthogonal Chebyshev moments. The discrete orthonormal Chebyshev moments of an order $p+q$, with size $N \times N$ for an image $f(x, y)$, are defined in [44];

$$T_{pq} = \sum_{x=0}^{N-1} \sum_{y=0}^{N-1} \hat{t}_p(x) \hat{t}_q(y) f(x, y) \quad p, q = 0, 1, 2 \dots N-1. \quad (3.7)$$

while the recurrence relation are

$$\hat{t}_p(x) = \alpha_1 x \hat{t}_{p-1}(x) + \alpha_2 \hat{t}_{p-1}(x) + \alpha_3 \hat{t}_{p-2}(x) \quad (3.8)$$

$$p = 2, 3 \dots N - 1; x = 0, 1 \dots N - 1$$

where

$$\alpha_1 = \frac{2}{p} \frac{\sqrt{4p^2 - 1}}{\sqrt{N^2 - p^2}} \quad (3.9)$$

$$\alpha_2 = \frac{1 - N}{p} \frac{\sqrt{4p^2 - 1}}{\sqrt{N^2 - p^2}} \quad (3.10)$$

$$\alpha_3 = \frac{p - 1}{p} \sqrt{\frac{2p + 1}{2p - 3}} \sqrt{\frac{N^2 - (p - 1)^2}{N^2 - p^2}} \quad (3.11)$$

and the initial conditions of the above recurrence relation are

$$\hat{t}_0(x) = \frac{1}{\sqrt{N}} \quad (3.12)$$

$$\hat{t}_1(x) = \frac{\sqrt{3}(2x + 1 - N)}{\sqrt{N(N^2 - 1)}} \quad (3.13)$$

Unfortunately, the original formulation of the discrete orthonormal Chebyshev moments is not directly useful in this study since it is not invariant under translation. This is because the orthonormal Chebyshev moments has a linear relationship with geometric moments which also can be used as a proof that it is not invariant under translation. To show the relationship between orthonormal Chebyshev moments and the geometric moments, the scaled Chebyhev polynomial $\hat{t}_p(x)$ can be expressed as a polynomial of x as shown below;

$$\hat{t}_p(x) = \frac{1}{\beta(p, N)} \sum_{k=0}^p B_{p,k} \langle x \rangle_k \quad (3.14)$$

where

$$\begin{aligned} \beta(p, N) &= \text{sqr}(2p)! \binom{N+p}{2p+1} \\ B_{p,k} &= (-1)^{p-k} \frac{p!}{k!} \binom{N-1-k}{p-k} \binom{p+k}{p} \end{aligned} \quad (3.15)$$

Note that $\binom{N+p}{2p+1}$ is a combination number such that $\binom{p}{q} = p!/q!(p-q)!$. According to [14], $\langle x \rangle_k$ can be expanded using the generating function for the Stirling numbers of the first kind as shown below

$$\langle x \rangle_k = \sum_{e=0}^k s(k, e) x^e \quad (3.16)$$

and satisfy the following recurrence relations :

$$\begin{aligned} s(k, e) &= s(k-1, e-1) - (k-1)s(k-1, e) \\ k &\geq 1, e \geq 1 \end{aligned} \quad (3.17)$$

with

$$s(k, 0) = s(0, e) = 0, k \geq 1, e \geq 1 \quad (3.18)$$

By using equation 3.14 above, the discrete orthonormal Chebyshev moments is able to be expressed as a linear combination of geometric moments with the same order moments as shown in the equation below

$$\begin{aligned} T_{pq} &= \sum_{x=0}^{N-1} \sum_{y=0}^{N-1} \hat{t}_p(x) \hat{t}_q(y) f(x, y) \\ &= \sum_{x=0}^{N-1} \sum_{y=0}^{N-1} \frac{1}{\beta(p, N)} \sum_{k=0}^p B_{p,k} \sum_{e=0}^k s(k, e) x^e \frac{1}{\beta(q, N)} \sum_{l=0}^q B_{q,l} \sum_{f=0}^l s(l, f) y^f f(x, y) \\ &= \frac{1}{\beta(p, N)} \frac{1}{\beta(q, N)} \sum_{x=0}^{N-1} \sum_{y=0}^{N-1} \sum_{k=0}^p B_{p,k} \sum_{e=0}^k s(k, e) x^e \sum_{l=0}^q B_{q,l} \sum_{f=0}^l s(l, f) y^f f(x, y) \\ &= \frac{1}{\beta(p, N) \beta(q, N)} \sum_{k=0}^p \sum_{l=0}^q B_{p,k} B_{q,l} \sum_{e=0}^k \sum_{f=0}^l s(k, e) s(l, f) \sum_{x=0}^{N-1} \sum_{y=0}^{N-1} x^e y^f f(x, y) \\ &= \frac{1}{\beta(p, N) \beta(q, N)} \sum_{k=0}^p B_{p,k} \sum_{l=0}^q B_{q,l} \sum_{e=0}^k s(k, e) \sum_{f=0}^l s(l, f) m_{ef} \end{aligned} \quad (3.19)$$

where m_{ef} is the geometric moments of order $e + f$, as defined below

$$m_{ef} = \sum_{x=0}^{N-1} \sum_{y=0}^{N-1} x^e y^f f(x, y) \quad (3.20)$$

3.3.2 Translation Invariant Orthonormal Chebyshev Moments

The previous section has shown that there is a linear relationship with orthonormal Chebyshev moments and geometric moments. However, it is known that the geometric moments are not invariant under translation. In order to provide translation invariance in the discrete orthonormal Chebyshev moments, each pixel will be subtracted with the centroid position so that it is independent of position. Therefore, the basis in geometric moments (x, y) will be replaced with the basis of central moments $(x - x_c, y - y_c)$ into equation 3.19, as below ;

$$\begin{aligned}
 C_{pq} &= \frac{1}{\beta(p, N)\beta(q, N)} \sum_{k=0}^p \sum_{l=0}^q B_{p,k} B_{q,l} \sum_{e=0}^k \sum_{f=0}^l s(k, e) s(l, f) \\
 &\quad \sum_{x=0}^{N-1} \sum_{y=0}^{N-1} (x - x_c)^e (y - y_c)^f f(x, y) \\
 &= \sum_{x=0}^{N-1} \sum_{y=0}^{N-1} \frac{1}{\beta(p, N)} \sum_{k=0}^p B_{p,k} \sum_{e=0}^k s(k, e) (x - x_c)^e \\
 &\quad \frac{1}{\beta(q, N)} \sum_{l=0}^q B_{q,l} \sum_{f=0}^l s(l, f) (y - y_c)^f f(x, y) \\
 &= \sum_{x=0}^{N-1} \sum_{y=0}^{N-1} \hat{t}_p(x - x_c) \hat{t}_q(y - y_c) f(x, y) \tag{3.21}
 \end{aligned}$$

Equation 3.21 has shown that the orthonormal Chebyshev moments has relationship with the central moments. This shows that the orthonormal Chebyshev moments also have the property of being invariant under translation. Hence, using equation 3.7, we replace each (x, y) by subtracting them with the centroid (x_c, y_c) . Here, the Translation Invariant Orthonormal Chebyshev Moments (henceforth, TIOCM) are defined by the following equation:

$$C_{pq} = \sum_{x=0}^{N-1} \sum_{y=0}^{N-1} \hat{t}_p(x - x_c) \hat{t}_q(y - y_c) f(x, y) \quad p, q = 0, 1, 2 \dots N - 1. \tag{3.22}$$

where

$$\begin{aligned}
 \hat{t}_p(x - x_c) &= \alpha_1(x - x_c) \hat{t}_{p-1}(x - x_c) + \alpha_2 \hat{t}_{p-1}(x - x_c) \\
 &\quad + \alpha_3 \hat{t}_{p-2}(x - x_c) \tag{3.23}
 \end{aligned}$$

$$p = 2, 3 \dots N - 1; x = 0, 1 \dots N - 1$$

and the remaining variables remain the same, except the initial conditions of the above recurrence relation are

$$\hat{t}_0(x - x_c) = \frac{1}{\sqrt{N}} \quad (3.24)$$

$$\hat{t}_1(x - x_c) = \frac{\sqrt{3}(2(x - x_c) + 1 - N)}{\sqrt{N(N^2 - 1)}} \quad (3.25)$$

and x_c and y_c are the centroids of object or objects in the image as described by equation 3.5.

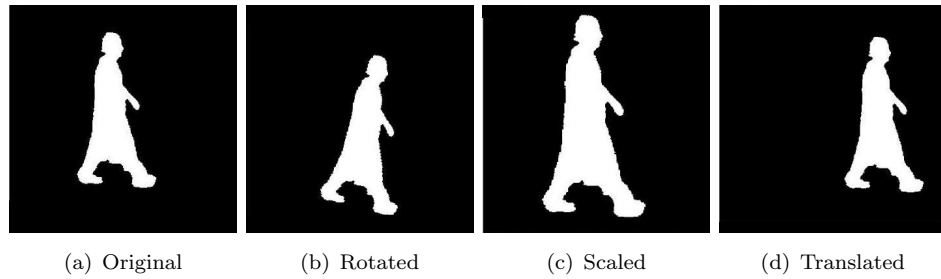


FIGURE 3.2: Image of a Silhouette (taken from Southampton Gait Database), (a)Original, (b)Rotated (c)Scaled (d)Translated

Here, we enclose a numerical experiment on comparing the original formula of Orthonormal Chebyshev Moments by Mukundan and our modified Orthonormal Chebyshev Moments. This numerical experiment shows that the modified Orthonormal Chebyshev moments are invariant under translation. The experiment is tested on an individual subject's silhouette as shown in Figure 3.2. The silhouette image was taken from the Southampton Gait Database and it has been translated and rotated for this comparison. Meanwhile the results are shown in Table 3.1 displaying a comparison between Orthogonal Chebyshev moments, Orthonormal Chebyshev moments, and the new method, the TIOCM, for the tested images. Compared to orthogonal and orthonormal Chebyshev moments, the TIOCM shows that moments of translated image are equal to moments of original image, proving that the new method is invariant under translation.

In addition, this formulation is similar and concurrent to work by [81]. This contribution was presented earlier as mentioned in section 1.3.2, while [81] presented a contribution which was published by the Journal of the Pattern Recognition in 2007. This indicates that this contribution is still new and very interesting.

Chebyshev Moments	Moments Order	Original Image C_{pq}	Rotated Image R_{pq}	Scaled Image S_{pq}	Translated Image T_{pq}
Orthogonal	00	0.09	0.09	0.16	0.09
	20	-0.22	-0.22	-0.36	-0.11
	02	-0.13	-0.08	-0.13	-0.13
	11	-0.00	0.00	0.00	0.06
Orthonormal	00	38.76	38.75	65.51	38.76
	20	46.78	45.95	80.97	66.90
	02	63.10	71.57	122.58	63.10
	11	-1.01	0.03	0.55	8.35
Orthonormal (Modified)	00	38.76	38.75	65.51	38.76
	20	174.99	175.07	298.37	174.99
	02	189.11	188.94	338.65	189.11
	11	116.49	114.25	197.74	116.49

TABLE 3.1: Comparison Chebyshev Moments for images of silhouette

3.4 Summary

Analyses on static images using moments have been presented in this chapter. A new method TIOCM was also proposed in this chapter. This method had a linear relationship with central moments, showing that it also has the property of invariant under translation. Later, by using real data, this method will be applied in chapter 5 for estimating crowd density. The experimental data for the estimation of crowd density will be described in the next chapter.

Chapter 4

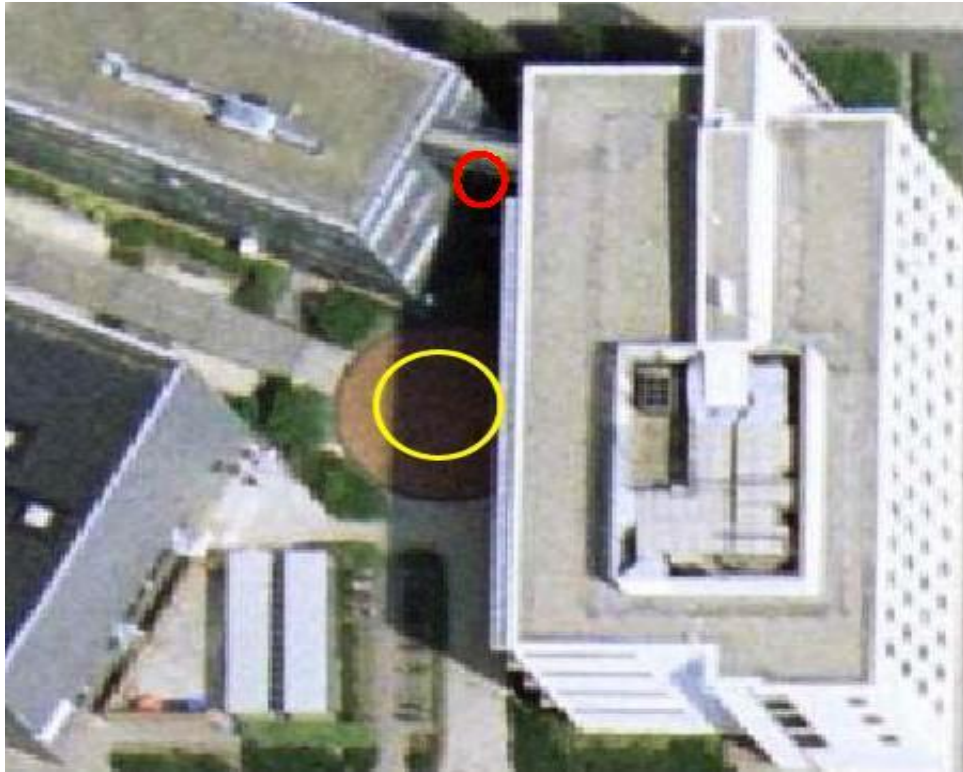
Experimental Data

4.1 Introduction

The first stage of monitoring crowds is to observe people through CCTV followed by obtaining images of crowds. However, the collection of real data may incur high cost and will become more difficult as it needs to deal with the organisation that controls the events and the venue where crowds may occur. As an alternative, images of crowds were taken during a graduation day, simulating a surveillance application.

This chapter describes the experimental data. The data used for this study is a scene of ECS Graduation Day, captured in front of the Zepler Building, University of Southampton. This scene was taken from 10am to 6pm with a break between 12 and 2 in the afternoon. It was taken in July 2005, when the weather was both sunny and cloudy. The view was captured from the highest level of a transparent corridor as shown in figure 4.1. Figure 4.1(a) displays the target location from top view, taken from Google Earth Software, while Figure 4.1(b) displays the target location from lateral view, taken from <http://www.ecs.soton.ac.uk/regenesis/pictures/>. The distance between the camera and the ground where the crowds appeared was approximately 20m.

The scene was captured using a Canon MV30i progressive scan miniDV camcorder and formatted as an avi file, then transferred into individual frames using VideoMach 2.7.2 [50]. The image shown in Figure 4.2, shows a view of people walking in and out of the food tent or the nearby building, people gathering, standing and talking to each other in front of the Zepler building and most of them were taking photos of each other. Each image was originally 720×576 pixels and was then cropped to 200×200 pixels. Figure 4.2 shows the original size of image and the grey level square illustrates the location to be used in the experiment. The area in the real world of the target image was approximately $13m^2$.



(a)Top view



(b)Lateral view

FIGURE 4.1: The red circle shows the location of the camera, while the yellow circle refers to the location of crowds taken from (a) the top and (b) lateral views



FIGURE 4.2: The location of the target area (shown in grey level square) in the original size of image.

4.2 Data Organization

This section describes the image data that was used for this study. The experimental data for crowd density are static images. Review of crowd analysis application on crowd density as shown in section 2.3.1, concludes that there are two different approaches to estimate crowd density, such as providing density in numbers or within a range of density classes. The latter approach was chosen as it is more appropriate. The Level of Service [42] was referred to provide the range of density, by making arrangement to suit the target area, and produced five range of class density from very low to very high. This is shown in Table 4.1. Examples of images according to the range of class density are also shown in Figure 4.3.

The images were recorded every 10 seconds, which appears a sufficient sampling rate to represent the underlying motion. For automated crowd density estimation, the ground truth of people in the images will be required. Here, the ground truth of the number

TABLE 4.1: Level of Service

<i>Level of Service</i>	<i>Range of Density(people/m²)</i>	<i>Range of People</i>	<i>Group</i>
A : Free(normal) flow	< 0.5	< 7	Very Low
B : Restricted flow	0.5 - 0.80	7 - 10	Low
C1 : Dense flow	0.81 - 1.26	11 - 16	Moderate
C2 : Very dense flow	1.27 - 2.0	17 - 26	High
D : No Flow(Jammed)	> 2.0	> 27	Very High

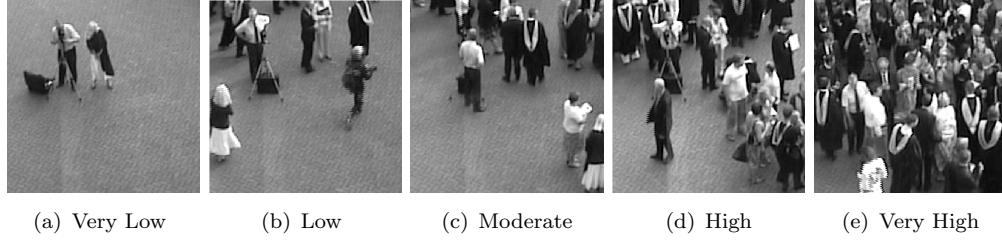


FIGURE 4.3: Sample Image of Crowds

of people in the crowd was counted manually with a condition that it was counted as a person if it showed a shape of the whole body, or a part of body such as head only, or combination of head and abdomen, or combination of waist and leg.

This experiment contains three different sets of image data, such as morning, afternoon and combined data. The morning data are images taken from 10am to 12pm, while the afternoon data are images taken from 2pm to 6pm. The combined data are combination image taken from morning and afternoon datasets. Each dataset comprises a range of density and is split into training and testing data, as described in Table 4.2. Meanwhile, sample images that were used in this experiment are given in Appendix A.

TABLE 4.2: Number of images used in Training and Testing Data

	Training Data						Testing Data					
Type	VL	L	M	H	VH	All	VL	L	M	H	VH	All
Morning	10	10	10	10	10	50	5	5	5	5	5	25
Afternoon	10	10	10	10	10	50	5	5	5	5	5	25
Combined	10	10	10	10	10	50	5	5	5	5	5	25

4.3 Boundary Binary Image

For the purpose of computational efficiency, a boundary binary image was considered. To simplify, these images were transformed from grey level to binary images. Rather than using a binary image where the objects were shaped like blobs, the objects were observed by their edges. This process was believed to be more efficient in storing data and time.

To obtain a boundary binary image, phase congruency was considered. Phase congruency is an edge detector proposed by [34], and was based at points in an image where the Fourier components were maximally in phase. Comparison of edge detectors between canny (as an optimal edge detector on spatial domain) and phase congruency, done by [34, 80], shows that the latter edge detector performs better.

Phase congruency has significant advantages as it is a dimensionless quantity that was invariant to changes in brightness, and allowed a universal threshold value that could be applied to wide range of images [34, 80]. As a result, the problem of selecting an optimum threshold for the experimental image data of crowds with varying of illumination due to weather may be solved.

In reference to [34], the phase congruency algorithm was applied. To obtain the binary image, a threshold based on the average pixel value of the phase congruency image was applied. Since the scene has a background of brick pavement, the results shown in Figure 4.4 produced some noise especially when there were few people in the scene. Then a 5×5 average filter was then used to smooth the image data, thus eliminating noise. Examples of the original image, phase congruency image and the boundary binary image are shown in Figures 4.4(a), 4.4(b) and 4.4(c) respectively. The features of boundary binary images were then used in measuring the density of crowd.

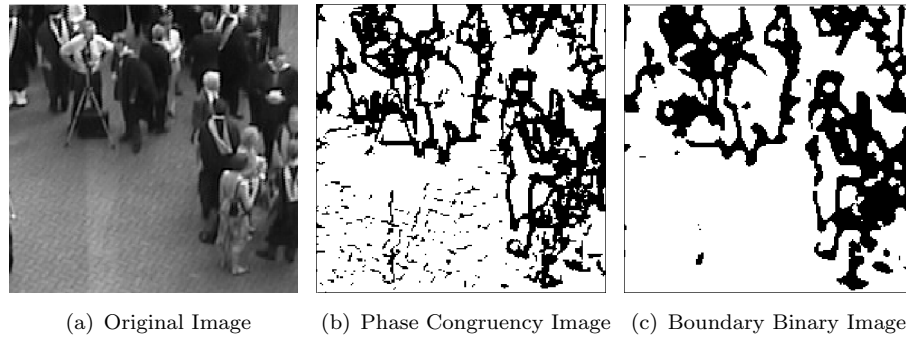


FIGURE 4.4: Sample Crowds Image for Boundary Binary Image

4.4 Summary

This chapter has presented a description of image data that were used for estimating crowd density and observing crowd movement. However, another measurement on the location target area of the crowd image was never be possible, due to the fire on Sunday October 30th 2005.

It was clear now that the target of this thesis is related to crowd density. Later in this report, moments were used to extract the features from the experimental data of still images. Meanwhile, the results are analysed and discussed in chapter [5](#).

Chapter 5

Measuring the Density of Crowds

The crowd loves density. It can never feel too dense. Nothing must stand between its parts or divide them; everything must be the crowd itself. The feeling of density is strongest in the moment of discharge. One day it may be possible to determine this density more accurately and even to measure it [11].

5.1 Introduction

The aim of this chapter is to evaluate the effectiveness of the new method called TIOCM. This method was used as a feature extractor where the problem is crowd density. The reason for estimating crowd density was described earlier in section 2.3.1. The experimental data was also described earlier in chapter 4.

Here, the first explanation in this chapter is an overview of the methodology of measuring the crowd density, in section 5.2. Three different techniques were used to extract features. They were Gray Level Dependency Matrix method (henceforth, GLDM), the Minkowski Fractal Dimension method (henceforth, MFD) and the TIOCM method. Explanations of their methodology and results have been written in section 5.2.1, 5.2.2 and 5.2.3 respectively. Since TIOCM are the main feature extractor, the other two techniques mentioned above were chosen to be compared and a comparison of results has been provided in section 5.2.4. Finally there is a further analysis in section 5.3.

5.2 Overview of Experimental Methodology

In this experimental study, there were three different methods used to extract features from images in order to estimate crowd density. The TIOCM was the new contribution method in this study, while the GLDM and MFD were known as the best previous methods for estimating crowd density from the work of [39, 42]. These three feature extractors were then compared to identify the best method with the best performance.

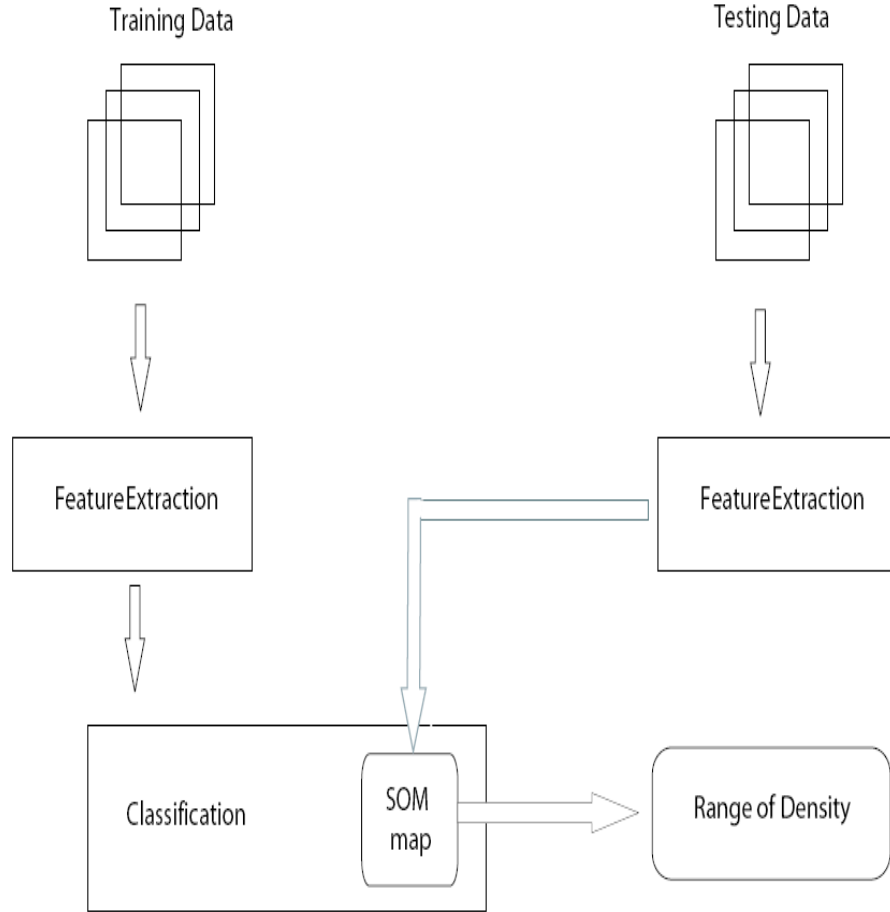


FIGURE 5.1: A General Process for the Methodology

Generally, these methods implement similar processes as shown in Figure 5.1. The first step took a selection of images which were divided into training and test data. These data were sent to the feature extractor. Next, the extracted features of the training data were submitted to a Self Organizing Map (SOM) classifier.

The SOM classifier was proposed by Kohonen [33] as a technique to aid visualization and interpretation of large and high-dimensional data sets, reducing them onto a much lower dimensional network in an orderly manner. [39, 42] used a SOM to classify the images of

crowd density within specified ranges, using it both to reduce the dimensionality of the GLDM and as a final classifier. We have chosen not to use other classifiers to maintain fidelity with his work.

The SOM Toolbox [69] was used in this study. SOM was also known as a supervised learning classification process where the aim to predict the class of a new instance from a sample of labelled examples. This supervised classification approach starts by building a classifier on training data and then implements the classifier for the test data. The training and test data were assigned according to a class label. The class label used in this experiment is a range of density from very low, low, moderate, high and very high density. Finally, to observe the accuracy of the classification performance, the class label from the manual process (ground truth) and from the automatically process (using SOM classification) were compared.

A brief explanation of SOM is described here since the three methods presented in this chapter used the same classifier in their classification process. A SOM contains a number of neurons which are represented by a d -dimensional weight vector $\mathbf{m} = [m_1, m_2 \dots m_d]$ where d is equal to the dimension of the input feature vector. First, the weight vectors are initialized with small random values. Then in each training step, a sample vector is chosen randomly from the input data \mathbf{x} and a similarity measurement between it and all weight vectors of the SOM map is calculated. The similarity measurement is usually defined by a distance measure such as Euclidean distance. The Best Matching Unit (BMU), denoted as c represents the greatest or closest similarity with the input sample \mathbf{x} and can be defined as below;

$$\|\mathbf{x} - m_c\| = \min_i \{\|\mathbf{x} - m_i\|\} \quad (5.1)$$

where $\|\cdot\|$ is the distance measure.

After the BMU was determined, the BMU and its neighbours were updated and moved towards the input vector in the input space according to the equation below;

$$m_i(t+1) = m_i(t) + \alpha(t)h_{ci}(t)[\mathbf{x}(t) - \mathbf{m}_i(t)] \quad (5.2)$$

where t denotes time, $\mathbf{x}(t)$ is an input vector taken randomly from the input data at time t , h_{ci} is the neighbourhood kernel around the BMU and $\alpha(t)$ as a learning rate is a decreasing function of time between $[0, 1]$. Here, the neighbourhood kernel around the BMU is the Gaussian neighbourhood function, defined as ;

$$h_{ci} = \exp\left(-\frac{\|r_c - r_i\|^2}{2\sigma^2(t)}\right) \quad (5.3)$$

where r_c is the location of unit c and r_i is the location of the neighbourhood node on the SOM map while σ^2 is the neighbourhood radius at time t . In early stages of training, relatively large initial learning rate α_0 and the neighbourhood radius σ^2 are used. As the training progresses, the neighbourhood radius is decreased with time. At the beginning of the training stages, SOM learns to roughly cover the space, while in the later stages, SOM fine tunes to describe the local details. After training, the SOM map is labelled according to the range of density and will then be used for labelling the test data.

5.2.1 Translation Invariant Orthonormal Chebyshev Moments

This section describes the method of TIOCM and its results. The first step is to process each image into a binary image as described in chapter 4. Then, six TIOCM are calculated for each image. The features are TIOCM with moments order $(p, q) = 00, 10, 01, 20, 02$. These feature vectors will be submitted to the SOM classifier to obtain a range of density from very low up to very high density.

In this study, moments can also be viewed as a texture descriptor. For example, the zero order moments can be seen to be like the area of the target object. A small area of edge points can be viewed as a coarse texture which represents low crowd density while a huge quantity of edge points can be viewed as a fine texture which indicates a high crowd density.

TABLE 5.1: Chebyshev Moments Results : Training and test data for morning (MD), afternoon (AD) and combined (CD) datasets

Datasets	Training Data						Testing Data					
	VL (10)	L (10)	M (10)	H (10)	VH (10)	Total (50)	VL (5)	L (5)	M (5)	H (5)	VH (5)	Total (25)
MD	9	9	9	8	7	42	2	4	3	3	5	17
AD	10	9	7	9	10	45	5	5	3	5	5	23
CD	9	8	8	9	8	42	4	3	4	5	5	21

Table 5.1 shows the quantity of images that have obtained correct classifications for training and testing of morning, afternoon and combined datasets. The correct classification is done by matching the range of density from SOM classification and the ground truth. If the range of density from the SOM classification is equal to the range of density from the ground truth, then the counting is considered correct.

The results shown in Table 5.1 were separated according to training and testing data, and into a range of density such as very low (VL), low (L), moderate (M), high (H) and very high (VH) density. For training data, out of 50 images of morning and combine datasets, only 42 images were correctly classified, while the afternoon datasets produced 45 correct images out of 50 images. Meanwhile, for testing data, out of 25 images, the afternoon datasets produce 23 correct classification images followed by combine datasets with 21 correct images and later 17 correct images from the morning dataset.

Overall, the afternoon dataset has produced better performance compared with morning and combined dataset, with 90% correct classification for training data and 92% classification for testing data. This happens as the afternoon dataset have less shadow compared to the morning and combined datasets. The training data of all datasets seem to produce good classification especially for very low density while the test data only perform 100% true classification for very high density in all types of datasets.

The fact that the training data did not get 100% in its classification performance is due to the existence of shadow and noise which these were also counted as person. Alternatively, to count each person with no shadow and noise using the image processing technique is a complex process, compared to counting each person in the image manually based on human perception. Thus, to improve the classification performance, shadow and noise should be eliminated as much as possible. By doing so, this may be able to improve the correct classification for both the training and test data.

5.2.2 Grey Level Dependency Matrix

This section presents the methodology and results of using GLDM as the feature extractor. The GLDM proposed by Haralick [26], has been used widely in various applications such as in satellite imagery, aerial, and microscopic imagery. GLDM is also known as spatial grey level dependency matrix, grey level co-occurrence matrix or gray tone dependency matrix [26, 39].

In general, GLDM can be considered as a second-order joint conditional probability density function, $f(i, j|d, \theta)$ which calculates the probability of the pair of grey levels (i, j) occurring in the image given, where these pixels are separated by a distance d and a direction θ . The pair of brightness levels (i, j) of the function $f(i, j|d, \theta)$ is defined by:

$$f(i, j|d, \theta) = \sum_{x=1}^N \sum_{y=1}^N (P_{x,y} = i) \wedge (P_{\hat{x},\hat{y}} = j) \quad (5.4)$$

where $P_{x,y}$ and $P_{\hat{x},\hat{y}}$ are of the same image, the x co-ordinate \hat{x} and the y co-ordinate \hat{y} are offset given by $\hat{x} = x + d\cos(\theta)$ and $\hat{y} = y + d\sin(\theta)$ respectively, with $d = 1$ and $\theta = 0^\circ, 45^\circ, 90^\circ, 135^\circ$. Four measurements namely: a) Contrast Ct , b) Homogeneity Ho , c) Energy En and d) Entropy Et were referred to [26, 39], with L as the number of grey levels of the image. These measurements describing the GLDM are shown in equations below.

Contrast:

$$Ct(d, \theta) = \sum_{i=0}^{L-1} \sum_{j=0}^{L-1} (i-j)^2 f(i, j|d, \theta)$$

Homogeneity:

$$Ho(d, \theta) = \sum_{i=0}^{L-1} \sum_{j=0}^{L-1} \frac{f(i, j|d, \theta)}{1 + (i-j)^2}$$

Energy:

$$En(d, \theta) = \sum_{i=0}^{L-1} \sum_{j=0}^{L-1} f(i, j|d, \theta)^2$$

Entropy:

$$Et(d, \theta) = - \sum_{i=0}^{L-1} \sum_{j=0}^{L-1} f(i, j|d, \theta) \log f(i, j|d, \theta)$$

For implementation, the GLDM method is applied to a number of images of an $N \times N$ size image. Sixteen features will be produced by the GLDM method for a moving tiled window of size $w \times w$ sub-images with an interval of $w/2$. In total, there will be $16 \times (2N/w - 1)^2$ features for each image. These feature vectors will be presented to a SOM classifier. In addition, there are two SOM classifiers used in this GLDM method. The first SOM is a part of the feature extraction process and the second SOM is the classifier used to classify features into range of density. The former SOM is an unsupervised learning process. The purpose of sending these feature vectors into the former SOM is to learn the features and group them according to similar features into a number of cluster, g . This experiment will be tested using $w = (20, 40, 50)$ and $g = (7, 14, 21, 28, 35, 42, 49)$, in order to seek optimum results of classification. Although the sub-images size could be $w = 10$, it was rejected as classification using features from $w \times w = 10 \times 10$ was time consuming.

The first SOM will label the sub-images into g groups of clusters and then generate a histogram for each image, as shown in Figure 5.2. Next, each histogram which represents each image will be an input feature to the second SOM to give a classification within the range of very low density up to very high density. The histogram with many

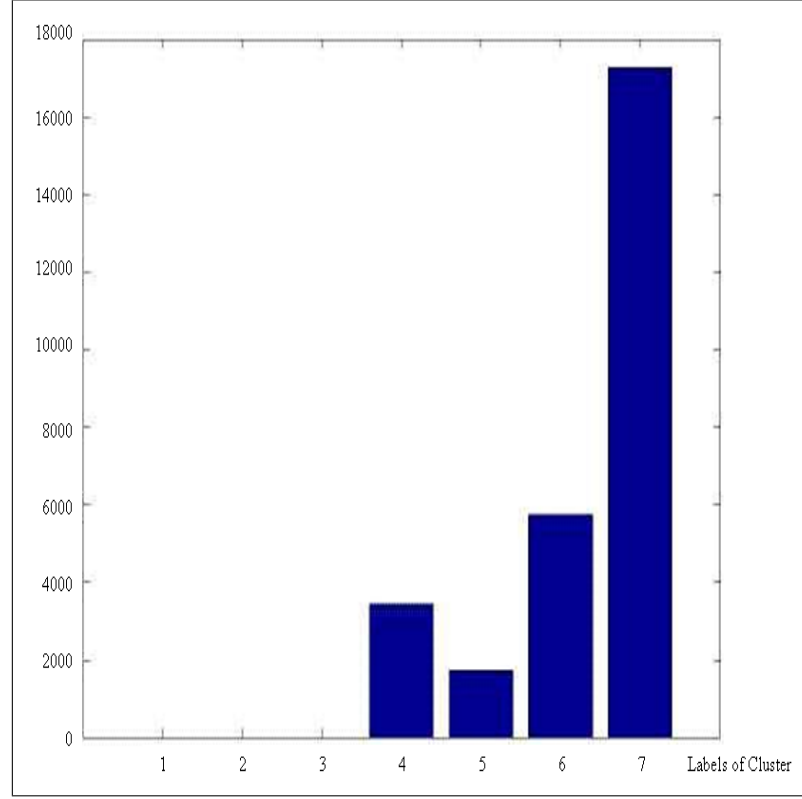


FIGURE 5.2: An example of a histogram image.

occurrences of low grey level indicates low occupation while many occurrences of high grey level indicates high occupation. The relationship of the histogram pattern and the level of occupation is learned by the second SOM classifier in the training stage. This information was then used to estimate the density for test data. Note that the GLDM experiment was used for three different datasets such as morning, afternoon and combined dataset.

TABLE 5.2: GLDM Results: Training data for Morning (MD), Afternoon (AD) and Combined (CD) datasets

Size Window	20			40			50		
Datasets	MD	AD	CD	MD	AD	CD	MD	AD	CD
Clusters									
7	68	82	78	80	92	70	78	96	78
14	78	86	90	72	90	74	72	90	86
21	78	94	86	76	86	88	74	90	82
28	82	90	88	76	84	76	80	92	86
35	84	94	82	78	86	76	76	88	86
42	80	92	82	74	80	78	80	86	84
49	80	88	80	72	82	80	76	86	82

Table 5.2 and 5.3 show the results of crowd density estimation obtained from the images of training and test data respectively for morning, afternoon and combined datasets.

TABLE 5.3: GLDM Results: Test data for Morning (MD), Afternoon (AD) and Combined (CD) datasets

Size Window	20			40			50		
Datasets	MD	AD	CD	MD	AD	CD	MD	AD	CD
Clusters	MD	AD	CD	MD	AD	CD	MD	AD	CD
7	64	72	78	72	84	60	60	72	68
14	68	88	90	80	80	72	64	92	76
21	68	88	86	60	84	72	64	88	76
28	64	92	88	72	72	68	72	76	68
35	72	76	82	68	72	52	68	88	76
42	76	76	82	68	88	68	72	84	76
49	80	88	80	76	73	64	68	88	72

The highest percentage classification for the morning dataset was 84% by using $(w, g) = (20, 35)$, while the highest percentage classification for the afternoon dataset was 96%, using $(w, g) = (50, 7)$ and the highest percentage classification for the combined dataset was 90%, using $(w, g) = (20, 14)$. Results from test data was based on the highest results obtained in the training data shown in Table 5.2. The optimum results of the test data, shown in Table 5.3, as according to the parameter (w, g) in the training data, were 72% for morning data, 72%, for afternoon data and 90% for the combined data. From these results, it seems that there are difficulties in obtaining an optimum performance classification for the test data based on only one size of window w and one number of clusters g .

Since the classification performance results does not achieve 100%, an investigation on the factors that lead to misclassification was done. It happens that during the feature extraction, there were shadows that were counted as one person. Consequently, the classification performance results did not achieve 100% in the training stage. It is suggested that, the shadow should be removed earlier before implementing the GLDM methods to obtain image features.

5.2.3 Minkowski Fractal Dimension

The methodology and results of the MFD technique are presented in this section. Generally, the fractal dimension is a measurement of roughness of a shape [21]. The advantage of choosing the MFD as the feature extractor is that it allows the estimation of the fractal dimension of a region and so can be used as fractal texture measure.

The Bouligand-Minkowsky fractal dimension [12] is defined from the Minkowsky sausages by analysing how the area of influence grows as the diameter D increases. The Minkowski

sausages method is a straightforward technique to calculate an area's influence by dilating a binary shape by a disk of diameter D [12]. For a single point the area of interest grows continually, however it tends to fill any holes in dense shapes so that it looks like a nearly filled region, growing more slowly. This concept is similar to the box-counting approach. The fractal dimension was obtained by analysing the log-log plot of the area of influence versus D , where curves with higher slopes were obtained for simple shapes. The Bouligand-Minkowsky [12, 42] fractal dimension is defined as $F = 2 - S$, where S is the slope of the log-log plot [42] defined by the logarithm of the number of white pixels, A , divided by the logarithm of the dilations size, r , as shown in equation 5.5.

$$S = \frac{\log(A)}{\log(r)} \quad (5.5)$$

The first step of this experiment was to obtain binary images from each input image. The process of transforming the original image into binary is described in section 4.3. The MFD was calculated for each binary image. Only one feature vector per image was generated in this experiment, and these features will be submitted to the SOM classifier to be assigned into the range of densities. In general, if the fractal dimension is equal to 2, it can be described as a region of square, which indicates the image full of crowds. In contrast, if the fractal dimension goes near to 0, it can be described as a dot, which indicates the image as a very low range of density.

Results using MFD are shown in table 5.4 and figure 5.3. Table 5.4 displays the quantity of correct classification images for training and testing data for all types of datasets. The table shows that out of 50 total training images, 28 images are correctly classified for afternoon datasets, followed by 23 and 22 correct training images for morning and combined data respectively. From 25 images of testing data in Table 5.4, the afternoon dataset obtained 60% correct images, followed by 48% correct images for morning dataset and 36% correct images for the combined dataset. In general, the very low density performed well in training and testing data for all datasets, but mostly the rest of the class density did not perform well.

TABLE 5.4: Minkowski Results : Training and test data for morning (MD), afternoon (AD) and combined (CD) datasets

Datasets	Training Data						Testing Data					
	VL (10)	L (10)	M (10)	H (10)	VH (10)	Total (50)	VL (5)	L (5)	M (5)	H (5)	VH (5)	Total (25)
MD	8	5	3	5	2	23	4	2	3	2	1	12
AD	10	7	3	1	7	28	5	4	1	2	3	15
CD	9	2	6	0	5	22	4	1	2	0	2	9

Most of the density classes were misclassified due to similar features obtained from the Minkowski experiment, which then seems to show difficulties in classification. This can be seen in Figure 5.3 where this figure plotted the fractal dimension and the number of people for the test data of all datasets to observe the relationship between them. From here, the fractal dimension was redundant due to the experiment of Minkowski that involved image dilation where this process had caused the noise and shadow to be dilated and then increased the area of noise and shadow to become similar to the target object. As a result, the MFD appears to lack discriminating capability with these images due to the process of image dilation, noise, and uncontrolled variation of illumination.

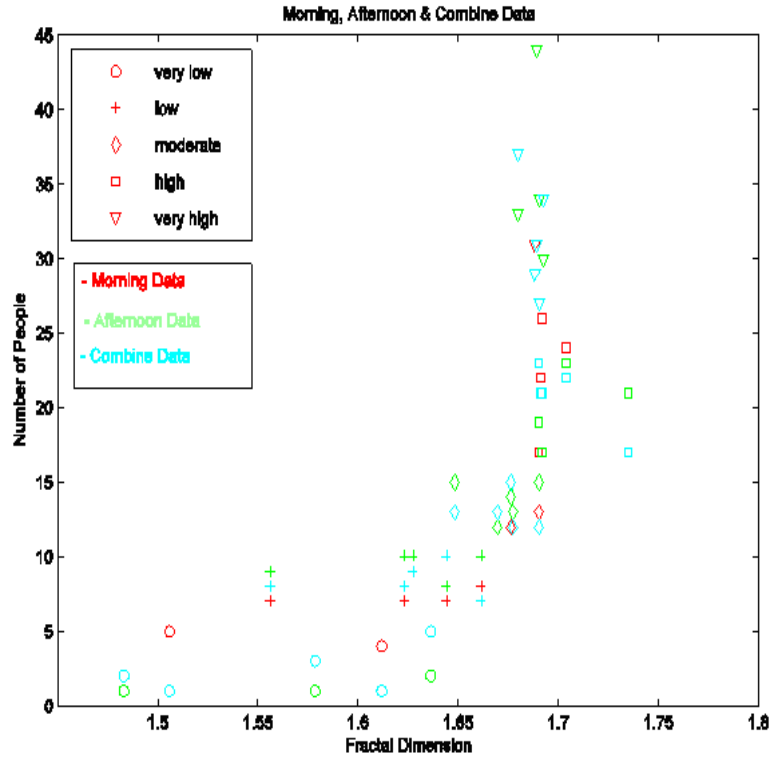


FIGURE 5.3: Fractal Dimension vs Number of People

Similar to the previous techniques, shadow and noise are the factors that influence the MFD algorithm to misclassify images. In addition, the process of image dilation to obtain the fractal dimension and also having the only one feature per image does not help to distinguish images into a suitable range of density. Since the MFD appears to lack discriminating capability with the input image, due to the misclassification factors as mentioned above, we conclude that this algorithm is not effective for estimating crowd density for outdoor scenes.

5.2.4 Comparison of Results

Since three different algorithms were used in this experiment, a comparison must be done to investigate the best performance in crowd density estimation. Figure 5.4 shows a graphical representation of the best results on test data, based on three different techniques according to three different datasets. Comparatively, MFD method has produced less than 60% classification performance, indicating the least effective method for estimating crowd density among the three techniques used in this study.

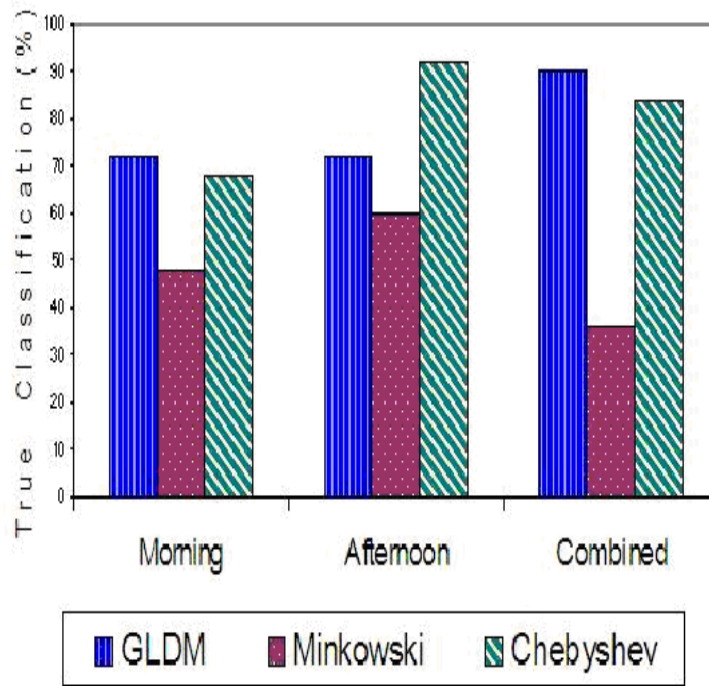


FIGURE 5.4: Comparison classification for test data of all datasets and all techniques

In Figure 5.4, the morning and afternoon datasets using GLDM methods arrived at the same percentage i.e 72%. However, the combined datasets rise much higher by achieving 90%. These results were based on the optimum percentage, performed by the training data as shown in Table 5.2, section 5.2.2. On the other hand, the training data was influenced by variation values of two parameters; the window size of sub-image, w and the number of clusters, g . As the optimum results of GLDM method was not based on the same value of w and g , thus, this shows that some more work needed to be done, in order to implement the GLDM method on different outdoor scenes of crowd images, using the same value of these two parameters.

Meanwhile, the TIOCM obtained 68% for morning, 92% for afternoon and 84% for combined data. Comparing these three datasets, all techniques performed best on the

afternoon dataset while the performance on the combined datasets was greater than the average of morning and afternoon datasets. This was caused by the small variation of illumination in the afternoon datasets compared to the morning datasets.

In general, the GLDM and the TIOCM, both outperform the MFD under all conditions. Comparatively, there was little to choose between the GLDM and the TIOCM, given the small number of samples in this experiment. However, for crowds in outdoor scene with small variations of illumination, the TIOCM method obtained 92% classification performance, proving that it performed better compared to 72% classification achieved by using GLDM method.

5.3 Further Analysis on Crowd Density Experiment

The previous section presented a comparison between three different methods which were used in this experiment. This comparison not only observed the performance of classification by the three methods but also according to the times of day which were morning, afternoon and combined datasets. In this section, a further analysis on this crowd density estimation experiment was done to observe on its statistical confidence. Using the same 75 images of morning and afternoon data image, each dataset was randomly divided to produce 5 observations, where each observation had 50 training data and 25 test data. These observations were not actually independent but were considered as independent.

Each observation was submitted to estimate crowd density using the three different methods: the GLDM, the MFD and the TIOCM. For each method, the training data was used to train a SOM with 5 clusters, corresponding to the number of density classes, which was subsequently used to classify the test data.

TABLE 5.5: Results of Training and Test Morning dataset using three different methods; TIOCM, MFD and GLDM

Type	Training Data					Testing Data				
	Ob1	Ob2	Ob3	Ob4	Ob5	Ob1	Ob2	Ob3	Ob4	Ob5
TIOCM	70	84	80	82	84	72	76	72	72	68
MFD	44	54	58	36	28	44	52	32	44	36
GLDM	74	90	84	80	68	64	72	60	68	64

Table 5.5 and Table 5.6 display results of correct classification in percentage for morning and afternoon dataset respectively. Generally, the results on the afternoon dataset gave better results compared to the morning session. This was because the afternoon dataset had a smaller variation of illumination compared to the morning data. The results in Table 5.5 and Table 5.6 show that both GLDM and TIOCM methods outperformed the

TABLE 5.6: Results of Training and Test Afternoon dataset using three different methods; TIOCM, MFD and GLDM

Type	Training Data					Testing Data				
	Ob1	Ob2	Ob3	Ob4	Ob5	Ob1	Ob2	Ob3	Ob4	Ob5
TIOCM	88	90	90	82	90	88	88	92	92	92
MFD	50	48	60	50	48	60	44	56	64	48
GLDM	88	92	92	72	78	68	92	84	80	78

MFD method, showing that on performance these results of experiments offer little to discriminate between them.

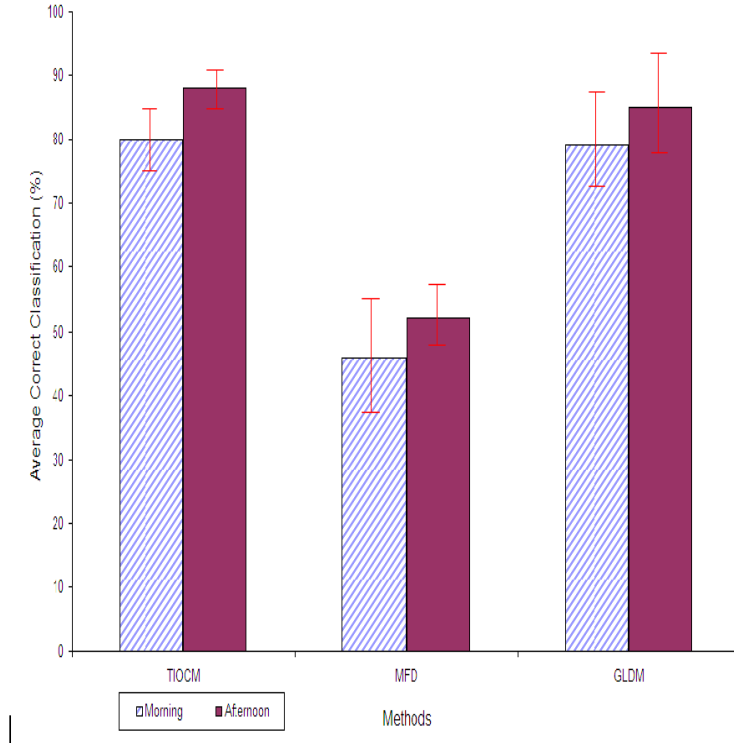


FIGURE 5.5: Comparison for Morning and Afternoon Training Data using three different methods ; TIOCM, MFD and GLDM

Meanwhile, Figure 5.5 and Figure 5.6 perform a comparison of measurement between the three methods in the morning and afternoon sessions using training and test data respectively. On average, TIOCM provide the highest correct classification for morning and afternoon datasets, followed by the GLDM method. The results of using TIOCM for morning or afternoon were close to its mean as the standard deviation was very small compared to other methods. The standard deviation for GLDM method was smaller than the MFD method in morning datasets but vice versa for afternoon dataset.

The TIOCM gave 75% to 85% correct classification for the training data and gives a range of 70% to 74% correct classification on test data. For the afternoon dataset, the

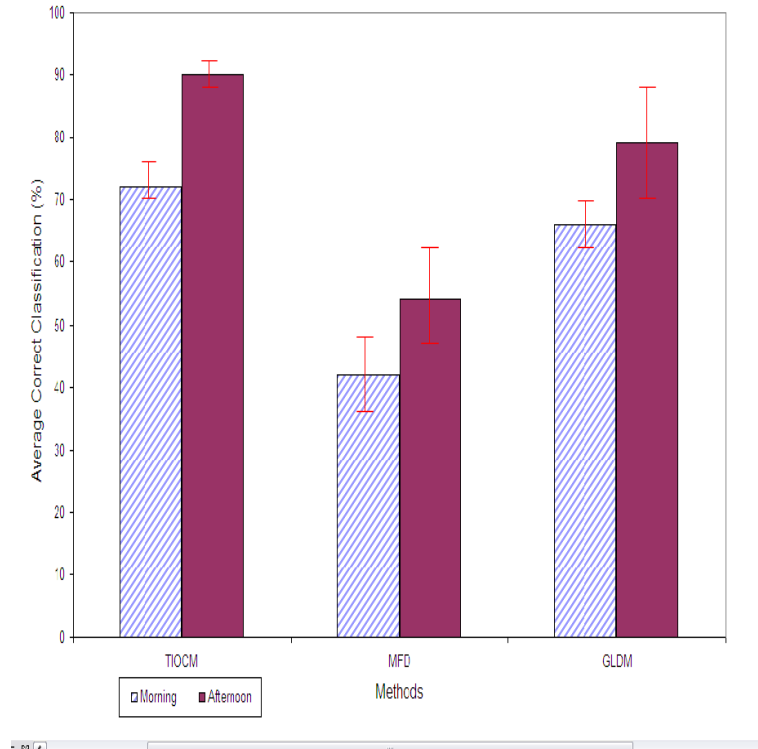


FIGURE 5.6: Comparison for Morning and Afternoon Testing Data using three different methods ; TIOCM, MFD and GLDM

correct classification for training data was in the range of 85% to 91 while the test data had a range of 88% to 92% of correct classification. In contrast, the MFD method gives less than 62% of correct classification for both morning and afternoon datasets.

Meanwhile, for morning session, the GLDM method performed 72% to 87% correct classification on the training data and 62% to 70% correct classification on test data. For afternoon dataset, the correct classification using the GLDM method for training data was in the range of 78% to 93%, while the test data had a range of 72% to 88% of correct classification.

Marana et. al (1997;1998) show that the GLDM is the best method for crowds at indoor scene performing 80% of correct classification while the MFD method obtained 78% correct classification. However, from the figure above, crowds at the outdoor scene showed that the TIOCM is as good as the GLDM, because the ranges of the confidence interval between both methods for morning or afternoon dataset mostly overlapped.

5.4 Summary

In conclusion, these results concur with [41], that for estimating crowd density in indoor scenes, the GLDM method perform better results as to compared with MFD. We then used these methods for outdoor scene and still the GLDM provided better results. Our new method, TIOCM, was also tested.

Generally, the GLDM and the TIOCM, both outperformed MFD under all conditions. Comparatively, there was little to choose between the GLDM and the TIOCM, given the small number of samples in this experiment. However, on both experiments shown in the previous section, the results using TIOCM method with condition that the images are crowds with small variation of illumination performed better compared to the classification results, achieved by GLDM method.

There was also some evidence that the GLDM required almost an order of magnitude more time to classify a test image, as compared with the TIOCM method. First, it is known that the GLDM method needs to calculate 16 features of $w \times w$ size sub-image for a $N \times N$ size of image. This is computationally expensive compared to TIOCM which only needed to calculate 6 features for an image. Further more, the GLDM needs to be classified in 2 SOM, where the first SOM learns the features of each sub-image and then classify the features into different number of cluster, while the second SOM takes the histogram of cluster image and classify them into a range of density. In contrast, TIOCM only uses one SOM classifier to label the image according to the range of density.

TABLE 5.7: Comparing performance between GLDM and Chebyshev

	<i>GLDM</i>	<i>Chebyshev</i>
Processor	Intel Pentium 4	Intel Pentium 4
CPU	1.70GHz	2.66 GHz
RAM	1.00 GB	0.99GB
Time	498s	15s

Based on figure 5.7, the GLDM method takes 498s to compute a single image from the early process of feature extraction until classification. In contrast, the TIOCM computes 33 times faster than the GLDM method. However, as both techniques were generated on different CPU, therefore, the performance may change slightly. Thus, if substantiated this is accepted, thus this indicates the TIOCM is more efficient and can be applied in practical situations.

Finally, it has been proved that moments can be used for groups of objects in which we used it for estimating crowd density. The TIOCM performed well in the afternoon datasets as it was influenced by small variation of illumination. Thus, we predict that the TIOCM may perform better in indoor scenes as compared with outdoor scene. The

following chapter is to investigate the usefulness of moments for measuring the movement of crowd.

Chapter 6

Conclusion and Future Work

In this chapter, the achievement of this thesis will be summarized. It begins with the motivation behind this work, followed by information related to crowd analysis and later the contributions and experiments on using moments in measuring the density and the movement of crowd. Finally, there were also some suggestion of possible ideas for future work to be expanded.

6.1 Conclusions

Crowds and public places are very difficult to protect. Especially, when a tragedy or danger strikes, it can often mean mass casualties, major damage as well as widespread panic and fear, which will influence loss of control, communication and containment. In a way to prevent this, surveillance application through monitoring crowds has been taken in full consideration. With the advanced technology, CCTV has been in use for monitoring crowds. Further more, with the integrating of computer vision techniques into the CCTV surveillance system, it has given an effective and affordable intelligent video in security systems. This integration of CCTV system and computer vision techniques may provide an automated crowd analysis which may also help to improve the prevention of incidents and accelerate action triggering in monitoring crowds.

Driven by the motivation of this study, we started with reviewing the related subject area that has interest with crowd. Along with this, we listed some terms that has been used in crowd analysis and provide a taxonomy of crowd analysis as written in Chapter 2. Relevant studies were not only considering crowds application related to computer vision but also based on a variety of disciplines which deal with crowds of people, such as psychology, sociology, safety and military.

Review on automated crowd analysis application shows that at least three important tasks should be considered in crowds application. They were related to crowd density, crowd movement and crowd behavior. In relation to the integration of the computer vision techniques and monitoring crowds application, this chapter 2 have listed several techniques of computer vision that have been implemented in these three crowds application. However, moments are lack in use for monitoring crowds. With this opportunity, moments were tested in measuring the crowd density.

Since moments have been chosen as the technique to be applied in this report, the following chapter has started by explaining the motivation of using moments and also describing a general information on using moments in vision application. Knowing that static images or still images were used in estimating crowd density and sequence of images were used in measuring crowd movement, chapter 3 provided analysis on static and sequence of images by using moment invariants. A new technique named as TIOCM was also introduced in this chapter. Despite that this technique has a linear relationship with the central moments, it also considers distinct information as they were orthogonal.

As the same real images of crowd were used in the experiment for crowd density and crowd movement, the explanation of these images have been written in Chapter 4. The data that were used in both experiment were outdoor scene images of crowds which were taken during a graduation day, simulating a surveillance application. In measuring the density of crowd, these images will be classified into five different ranges of density : very low, low, moderate, high and very high density.

In general, the focus of this study is to investigate the moment invariants as one of the computer vision techniques that can be applied in monitoring crowds application such as estimating crowd density. After doing the analysis using moments on static images as written in chapter 3, estimating crowd density using real images of crowds were tested out and the results were written in chapter 5.

In chapter 5, the new technique TIOCM was used for estimating crowd density. The technique was used as feature extractor and also compared to another two techniques named as GLDM and MFD. Static images were used as the input data and the results were then images labelled into 5 ranges of densities. Comparison on the performance were done base on these three feature extractor and also considering daytime. Above all, our new technique performed as good as the GLDM. However, if considering images with less shadow and time consuming, our new technique performed slightly better than the GLDM.

6.2 Future Work

In finishing this report, a number of suggested opportunities for further study appear and is presented in this thesis. We divide them into three categories: data; moments; and application of crowd analysis.

(i) Input Data

The input data used in this thesis were only images taken from one location area (the Zepler Building, University of Southampton). As the TIOCM has performed well in estimating crowd density, it will be much interesting to test this technique on a larger database. The database should consider different places, different weather, and including different timing.

(ii) Binary Image

Since using the image of crowd at an outdoor scene, the phase congruency technique was chosen in order to get features that appear to be invariant under illumination. However, the effect from the background surface that cause a lot of noise shows that the phase congruency technique is not very practical for images at outdoor scene. Therefore there is a need to do more research on selecting an appropriate technique to transform these sort of images to binary image with clear silhouette.

(iii) Moments

It is known that crowds can be viewed using microscopic approach or macroscopic approach. On another hand, in most computer vision's application, moments have done very well upon single object. The evaluation on a single object is viewed from a microscopic approach. Meanwhile, this study has been observing crowds from a macroscopic approach using moments. Hence, if moments could be viewed from both microscopic or macroscopic approach, then it could be related to the concept of superposition. If this could be proven, it may provide a much challenging and interesting research to be done.

(iii) Application of Crowd Analysis

- In order to investigate moments for crowd behavior, semantic description for crowd behavior based on interpreting low level information is required. There is a possibility that this future work can be divided into macroscopic and microscopic approach. The macroscopic approach shall focus on the pattern of crowd behavior, such as observing the normal and abnormal behavior of crowd. Meanwhile, the microscopic approach may focus on the behavior of

each individual person in the crowd. [57] is an example of using moments for recognizing individual activity.

- In addition, if there exist a relationship between moments of each individual person in the crowd and the moments of a crowd, then this may gain some interest to investigate the behavior of individual person in depth. Such an example is by applying image reconstruction to track the missing information about the target subject, and the Chebyshev moments have been successfully used for image reconstruction [44, 45].

Appendix A

Sample Images For Measuring Crowd Density

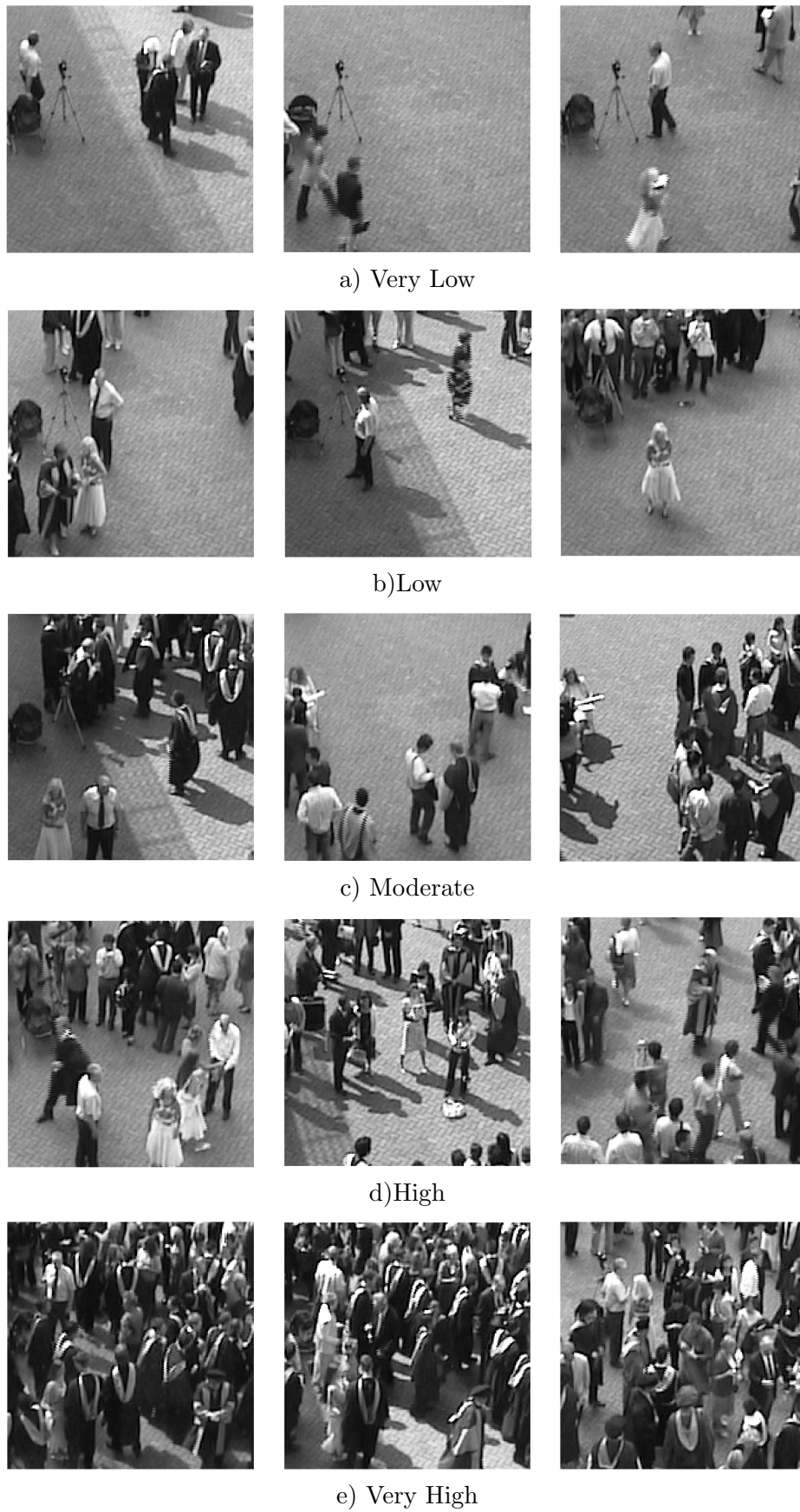


FIGURE A.1: Example of Training Data : Morning Session



FIGURE A.2: Example of Training Data : Afternoon Session



FIGURE A.3: Example of Testing Data : Morning Session



FIGURE A.4: Example of Testing Data : Afternoon Session

References

- [1] S. Al-nasur and P. Kachroo. A microscopic-to-macroscopic crowd dynamic model. In *Proceedings of the IEEE ITSC*, pages 606–611, September 2006.
- [2] S.J Al-nasur. *New Models for Crowd Dynamics and Control*. PhD thesis, Electrical and Computer Engineering, Virginia Polytechnic Institute and State University, December 2006.
- [3] Ernesto L. Andrade, Scott Blunsden, and Robert B. Fisher. Modelling crowd scenes for event detection. *icpr*, 1:175–178, 2006. ISSN 1051-4651. doi: <http://doi.ieeecomputersociety.org/10.1109/ICPR.2006.806>.
- [4] S. Y. A. Au, M. C. Ryan, M. S. Carey, and S. P. Whalley. Managing crowd safety in public venues: A study to generate guidance for venue owners and enforcing authority inspectors. Technical report, Health and Safety Executive, 1993.
- [5] Victor J. Blue and Jeffrey L. Adler. Modeling four directional pedestrian movements. Transportation Research Board 79th Annual Meeting, January 2000.
- [6] B.A Boghossian and J. Black. The challenges of robust 24/7 video surveillance systems. In *The IEE International Symposium on Imaging for Crime Detection and Prevention*, pages 33– 38, 2005.
- [7] B.A Boghossian and S.A Velastin. Real-time motion detection of crowds in video signals. In *IEE Colloquium on High Performance Architectures for Real-Time Image Processing*, volume 12, pages 1–6, 1998.
- [8] S. Bouchafa, D. Aubert, L. Beheim, and A. Sadji. Automatic counterflow detection in subway corridors by image processing. In *Journal of Intelligent Transportation Systems*, volume 6, 2001.
- [9] Robin Brant. Malaysia police break up protest. <http://news.bbc.co.uk/1/hi/7088877.stm?lsm>, November 2007.
- [10] A. Broggi, M. Bertozzi, A. Fascioli, and M. Sechi. Shape-based pedestrian detection. In *In Procs. IEEE Intelligent Vehicles Symposium 2000*, pages 215–220, 2000.

-
- [11] E. Canetti. *Crowds and Power*. Gollancz, 1962.
 - [12] R.M Cesar and L.D.F Costa. *Shape Analysis and Classification, Theory and Practice*. CRC Press, 2000.
 - [13] Christophe Chalons. Numerical approximation of a macroscopic model of pedestrian flows. *Journal of Scientific Computing*, 2005.
 - [14] Ch. A. Charalambides and J. Singh. A review of the stirling numbers, their generalizations and statistical applications. *Communications in Statistics-Theory and Methods*, 34(8):2533–2595, 1988.
 - [15] Fabio Chisari. 'the cursed cup': Italian responses to the 1985 heysel disaster. In *Source: Soccer and Society*, volume 5, pages 201–218(18). Routledge, part of the Taylor & Francis Group, 2004.
 - [16] N. Courty and T. Corpetti. Data-driven animation of crowds. In *Proceedings of Mirage 2007 - Computer Vision / Computer Graphics Collaboration Techniques and Applications*, Paris, France, March 2007.
 - [17] A. C. Davies, J. H. Yin, and S. A. Velastin. Crowd monitoring using image processing. *Electronics and Communication Engineering Journal*, 7(1), 1995.
 - [18] Google Earth. <http://earth.google.com/>, 2005.
 - [19] Michael Eysenck. *Psychology: An International Perspective*. Psychology Press, 2004.
 - [20] Z. Fang, S. M. Lo, and J. A. Lu. On the relationship between crowd density and movement velocity. *Fire Safety Journal*, 38(3):271–284, 2003.
 - [21] R. Fisher, K. Dawson-Howe, A. Fitzgibbon, C. Robertson, and E. Trucco. *Dictionary of Computer Vision and Image Processing*. Jon Wiley & Son Ltd, 2005.
 - [22] J. J. Fruin. The causes and prevention of crowd disasters. *Elsevier Science Publishers B.B.*, 1993.
 - [23] Anthony Giddens. *Social theory and modern sociology*. Stanford, Calif: Stanford University Press, 1987.
 - [24] Comp. Global Security. Field manual no. 19-15, civil disturbance, headquarters department of the army, washington. <http://www.globalsecurity.org/military/library/policy/army/fm/19-15/CH2.htm>, November 1985.
 - [25] Daniel J. Graham and David J. Field. Statistical regularities of art images and natural scenes: spectra, sparseness and nonlinearities. *Spatial Vision*, 21(1-2): 149–164, 2007.

-
- [26] R.M. Haralick. Statistical and structural approaches to texture. In *Proceedings of the IEEE*, volume 67, 1979.
- [27] I. Haritaoglu, D. Harwood, and L. Davis. W4 : Real-time surveillance of people and their activities. *IEEE Transactions On Pattern Recognition and Machine Intelligence*, 22(8), 2000.
- [28] Dirk Helbing, Lubos Buzna, Anders Johansson, and Torsten Werner. Self-organized pedestrian crowd dynamics: Experiments, simulations, and design solutions. *Transportation Science*, 39(1):1–24, 2005. ISSN 1526-5447. doi: <http://dx.doi.org/10.1287/trsc.1040.0108>.
- [29] M.K. Hu. Visual pattern recognition by moment invariant. *IRE Transactions Information Theory*, 8:179–187, 1962.
- [30] R. L. Hughes. The flow of human crowds. *Annu. Rev. Fluid Mech*, 35:169–182, 2003.
- [31] J. Kang, I. Cohen, and G. Medioni. Continuous tracking within and across camera streams. In *Proceedings of the IEEE Computer Vision and Pattern Recognition*, volume 1, pages 267–272, 2003.
- [32] Hubert Klüpfel. *A Cellular Automaton Model for Crowd Movement and Egress Simulation*. PhD thesis, University Duisburg–Essen, July 2003.
- [33] T. Kohonen. The self organizing map. *Proceedings of IEEE 1990*, 78(9):1464–1480, 1990.
- [34] Peter Kovesi. Image features from phase congruency. *Videre: A Journal of Computer Vision Research*, 1(3), 1999.
- [35] G. Le Bon. *The Crowd: A Study of the Popular Mind*. T. Fisher Unwin, 1931.
- [36] S. F. Lin, J. Y. Chen, and H. X. Chao. Estimation of number of people in crowded scenes using perspective transformation. In *IEEE Transactions on Systems, Man, and Cybernetics - Part A : systems and Humans*, 2001.
- [37] R. Ma, L. Li, W. Huang, and Q. Tian. On pixel count based crowd density estimation for visual surveillance. *IEEE Conference Cybernetics and Intelligent Systems*, 1, 2004.
- [38] John J Macionis. *Sociology*. Prentice Hall, 2005.
- [39] A. N. Marana, S. A. Velastin, L. F. Costa, and R. A. Lotufo. Estimation of crowd density using image processing. In *IEE Colloquium on Image Processing for Security Applications (Digest No: 1997/074)*, volume 11, pages 1–8, 1997.

- [40] A. N. Marana, S. A. Velastin, L. F. Costa, and R. A. Lotufo. Automatic estimation of crowd density using texture. In *Safety Science*, pages 165–175, 1998.
- [41] A. N. Marana, S. A. Velastin, L. F. Costa, and R. A. Lotufo. On the efficacy of texture analysis for crowd monitoring. In *Anais do XI SIBGRAPI*, 1998.
- [42] A. N. Marana, S. A. Velastin, L. F. Costa, and R. A. Lotufo. Estimation crowd density with minkowski fractal dimension. *Proceedings of IEEE*, pages 3521–3524, 1999.
- [43] R. Mukundan. Discrete orthogonal moment features using chebyshev polynomials. In *Proc. of International. Conference on Image and Vision Computing - IVCNZ'00*, pages 20–25, November 2000.
- [44] R. Mukundan. Improving image reconstruction accuracy using discrete orthonormal moments. In *Proc. of International. Conference On Imaging Systems, Science and Technology - CISST'2003*, pages 287–293, June 2003.
- [45] R. Mukundan. Some computational aspects of discrete orthonormal moments. *IEEE Transactions on Image Processing*, 13(8):1055–1059, 2004.
- [46] R. Mukundan and K.R. Ramakrishnan. *Moment Functions in Image Analysis- Theory and Applications*. World Scientific Publishing Co. Pte Ltd, Singapore, September 1998.
- [47] R. Mukundan, S.H. Ong, and P.A. Lee. Image analysis by tchebichef moments. *IEEE Transactions on Image Processing*, 10(9):1357–1364, 2001.
- [48] Soraia Raupp Musse, Branislav Ulicny, Amaury Aubel, and Daniel Thalmann. Groups and crowd simulation. In *SIGGRAPH '05: ACM SIGGRAPH 2005 Courses*, page 2, New York, NY, USA, 2005. ACM Press.
- [49] C. E. Nicholson. The investigation of the hillsborough disaster by health and safety executive. *Safety Science*, 18(4):249–259, 1995.
- [50] Z. Nikolic and R. Tibljan. www.gromada.com. web, 2003.
- [51] M. S. Nixon and A. Aguado. *Feature Extraction and Image Processing*. Newnes, 2002.
- [52] S. P. Prismall, M. S. Nixon, and J. N. Carter. On moving object reconstruction by moments. In *In Proceedings of British Machine Vision Conference*, pages 1–10, 2002.

- [53] S.P. Prismall. *Object reconstruction by moments extended to moving object sequences*. PhD thesis, Department Electronic and Computer Science, University of Southampton, 2005.
- [54] Hidayah Rahmalan, Mark S. Nixon, and John N. Carter. On crowd density estimation for surveillance. In *International Conference on Crime Detection and Prevention*, 2006. URL <http://eprints.ecs.soton.ac.uk/12852/>.
- [55] S. D. Reicher. The battle of westminster: developing the social identity model of crowd behaviour in order to explain the initiation and development of collective conflict. *European Journal of Social Psychology*, 26(1):115–134, 1996.
- [56] Stephen Reicher, Clifford Stott, Patrick Cronin, and Otto Adang. An integrated approach to crowd psychology and public order policing. *An International Journal of Police Strategies and Management*, 27(4):558–572(15), 2004.
- [57] R. Rosales. Recognition of human action using moment-based features. Technical report, Boston University Computer Science, 1998.
- [58] J. D. Shutler. *Velocity Moments For Holistic Shape Description of Temporal Features*. PhD thesis, Department Electronic and Computer Science, University of Southampton, 2001.
- [59] N.T. Siebel and S.J. Maybank. The advisor visual surveillance system. In *Applications of Computer Vision*, 2004.
- [60] J. D. Sime. Crowd psychology and engineering. *Safety Science*, 21:1–14, 1995.
- [61] G.K. Still. *Crowd Dynamics*. PhD thesis, Mathematics Department, Warwick University, August 2000.
- [62] Muralidhara Subbarao. *Interpretation of visual motion: a computational study, Series: Research Notes in Artificial Intelligence*. Pitman Publishers, London (distributed by Morgan Kaufman Publishers in USA), 1988.
- [63] Chua Sue-Ann. No abcs of protest reporting. <http://bersih.org/?p=516>, November 2007.
- [64] M.R. Teague. Image analysis via the general theory of moments. *Journal of the Optical Society of America*, 70:920930, August 1980.
- [65] Cho-Huak Teh and Roland T. Chin. On image analysis by the methods of moments. *IEEE Transactions Pattern Anal. Mach. Intell.*, 10(4):496–513, 1988. ISSN 0162-8828.

- [66] M. Tuceryan. Moment based texture segmentation. *Pattern Recognition letters*, 15: 659–668, July 1994.
- [67] N. Vaswani, A. R. Chowdhury, and R. Chellappa. Activity recognition using the dynamics of the configuration of interacting objects. *IEEE Computer Vision and Pattern Recognition*, 2003.
- [68] S.A. Velastin, J.H. Yin, A.C Davies, M.A. Vicencio-Silva, R.E. Allsop, and A. Penn. Analysis of crowd movements and densities in built-up environments using image processing. In *Image Processing for Transport Applications, IEE Colloquium*, pages 8/1 – 8/6, 1993.
- [69] J. Vesanto, J. Himberg, E. Alhoniemi, and P. Parhankangas. Report a57: Som toolbox for matlab 5. Technical report, Helsinki University of Technology, Finland, 2000.
- [70] R. Vincent and B Serge. Counting crowded moving objects. In *CVPR '06: Proceedings of the 2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pages 705–711, Washington, DC, USA, 2006. IEEE Computer Society. ISBN 0-7695-2597-0.
- [71] L. Wang and G Healey. Using zernike moments for the illumination and geometry invariant classification of multispectral texture. In *IEEE Transactions on Image Processing*, volume 7, pages 196 – 203, 1998.
- [72] E. W. Weisstein. Kullback-leibler distance from mathworld—a wolfram web resource. <http://mathworld.wolfram.com/Kullback-LeiblerDistance.html>, July 2004.
- [73] Dong Xu and Hua Li. Geometric moment invariants. *Pattern Recognition*, 41(1): 240–249, 2008. doi: doi:10.1016/j.patcog.2007.05.001.
- [74] D.B. Yang, H.H. Gonzalez-Banos, and L.J. Guibas. Counting people in crowds with a real-time network of simple image sensors. *Proceedings. Ninth IEEE International Conference on Computer Vision*, 1, 2003.
- [75] J.H. Yin, S.A. Velastin, and A.C Davies. Image processing techniques for crowd density estimation using a reference image. In *Second Asian Conference on computer Vision (ACCV95)*, volume III, pages 6–10, 1995.
- [76] T. Yusukey and N. Takashi. Scaling behavior of crowd flow outside a hall. *Physica A*, 292:545–554, 2001.
- [77] Beibei Zhan, Dorothy N. Monekosso, Paolo Remagnino, Sergio A. Velastin, and Li-Qun Xu. Crowd analysis: a survey. *Mach. Vision Appl.*, 19(5-6):345–357, 2008. ISSN 0932-8092. doi: <http://dx.doi.org/10.1007/s00138-008-0132-4>.

-
- [78] J. Zhang and T. Tan. Brief review of invariant texture analysis methods. *Pattern Recognition*, 35(3):735–747, 2002.
- [79] Li Zhang, Gong bin Qian, Wei wei Xiao, and Zhen Ji. Geometric invariant blind image watermarking by invariant tchebichef moments. *Optics Express Journal*, 15: 2251–2261, 2007.
- [80] Y. Zheng. *Automated Segmentation of Lumbar Vertebrae for the Measurement of Spine Kinematics*. PhD thesis, Department Electronic and Computer Science, University of Southampton, 2003.
- [81] Hongqing Zhu, Huazhong Shu, Ting Xia, Limin Luo, and Jean Louis Coatrieux. Translation and scale invariants of tchebichef moments. *Pattern Recognition*, 40(9): 2530–2542, 2007. ISSN 0031-3203. doi: <http://dx.doi.org/10.1016/j.patcog.2006.12.003>.