國 立 交 通 大 學

電子工程學系 電子研究所碩士班

碩 士 論 文

藉由模擬行人行為之不尋常行為偵測

Unusual Behavior Detection Based on Pedestrian Behavior Modeling

研 究 生：鄭郁霖

指導教授：王聖智 博士

中 華 民 國 一 百 年 九 月

藉由模擬行人行為之不尋常行為偵測

Unusual Behavior Detection Based on Pedestrian Behavior Modeling

研 究 生：鄭郁霖　　Student：Yu-Lin Cheng

指導教授：王聖智博士　　Advisor：Dr. Sheng-Jyh Wang

國 立 交 通 大 學

電子工程學系 電子研究所碩士班

碩 士 論 文

A Thesis

Submitted to Department of Electronics Engineering & Institute of Electronics

College of Electrical and Computer Engineering

National Chiao Tung University

in partial Fulfillment of the Requirements

for the Degree of Master

in

Electronics Engineering
September 2011
Hsinchu, Taiwan, Republic of China

中華民國一百年九月

藉由模擬行人行為之不尋常行為偵測

研究生：鄭郁霖　　　指導教授：王聖智 博士

國立交通大學

電子工程學系　電子研究所碩士班

摘要

在本論文中，我們提出一套自動化的行人異常行為偵測之監控系統。不同於一般監控系統是以環境場景分析的觀點出發來偵測異常行為的發生，我們是採取以行人的觀點為我們的出發點。行人的行為有千奇百種，若要仔細的去一一描述各種不同的行為是有困難的。但在一般的情況下，行人的行為是有跡可循的，並非如一般想像的隨機變動；亦即行人的行為都會因當時的周遭環境而採取在當下最適當的應對行為。因此描述周遭環境因素對行人行為所造成影響，將有助於我們去模擬各種行人的行為。在此，我們藉由行人的特性來描述行人的行為與周遭環境因素之間的關係，這些行人的特性包含了：目的地、舒適距離和朋友關係等。之後再去描述這些不同的特性對於行為所造成影響，來幫助我們來預測行人可能的行為模式進而幫助我們達到偵測異常行為發生的目的。相較於去建立所謂的場景模型，透過以行人的觀點來偵測異常行為的發生，我們將可以偵測出更多種抽象的異常行為發生。

Unusual Behavior Detection Based on Pedestrian Behavior Modeling

Student: Yu-Lin Cheng Advisor: Dr. Sheng-Jyh Wang

Department of Electronics Engineering, Institute of Electronics

National Chiao Tung University

Abstract

In this paper, we propose a surveillance system which is able to detect abnormal behaviors automatically. Unlike other approaches, we do not detect abnormal behaviors by constructing a scene model with normal behaviors. Instead, we emphasize the important aspect of pedestrian since it is the pedestrian who performs abnormal behaviors. There are actually various pedestrian behaviors and it is difficult to express every pedestrian behavior in detail. However, it is still possible to model pedestrian behavior since pedestrian behavior, in general, always reflects some kinds of close relationship with the current environment. Hence, we first take some social factors of the environment into account for the estimation of pedestrian properties. After that, we predict the possible pedestrian behaviors by using these estimated pedestrian properties. An abnormal behavior can be detected as long as the behavior of pedestrian is very different from the predicted behaviors. With this system, we are capable of detecting more abstract abnormal behaviors than before.

誌謝

在此要特別感謝我的指導教授 王聖智老師，在與您的學習中，讓我學習到了許多研究的方法與態度，也學習到了許多待人處事的技巧，雖然無法做到和老師您一樣，但卻使我在這一年中有著長足的成長。此外也要感謝實驗室的每個成員：感謝敬群、慈澄、禎宇、家豪學長們平時對我的指導與幫助，使我重新反省人生的目的；也要感謝韋弘、開暘、玉書、鄭綱平時之間的幫忙與鼓勵，使我在做研究的同時，添增了許多樂趣；也感謝秉修、彥廷、心憫、柏翔，使實驗室充滿了話題與歡笑。最後，我要感謝我的家人，以及所有曾經幫助過我的人，有你們的幫忙，我才能毫無後顧之憂的專心做研究，使我有今日的結果，再次謝謝大家。

Content

List of Figures

List of Tables

Chapter 1.

INTRODUCTION

As the rapid progress of computer industry, automatically visual surveillance systems have attracted much more attentions of researches than before. One of the reasons of the importance of visual surveillance systems is that once an intelligent surveillance system is developed, we can pay precious attentions only to those crucial events and objects. As a result, automatic visual surveillance has been applied widely in daily life. For example, analyzing traffic flows at intersection and inspecting luggage at airport. Although there are various applications of visual surveillance systems, the ultimate goal of researchers is the same. That is to develop a system which can detect abnormality automatically with high accuracy and can provide warnings to operators.

Due to the versatile applications of surveillance systems, it is almost impossible or at least very difficult to develop a generic surveillance system which can be applied to all different fields. As a result, in this thesis, we will only focus on those surveillance videos captured in non-crowded and wide-range scenes, as shown below in Figure 1-1.

Based on existing computer vision algorithms, it is still hard to analyze the body motion of pedestrians. Hence, most surveillance systems for non-crowded and wide-range scenes detect abnormal behaviors based on pre-trained scene models that are pre-learned from videos with normal-behavior pedestrians. Given a video, as long as the detected object in the video does not obey this scene model, we assume there is an occurrence of abnormal event. These approaches are mainly from the aspect of scene. However, it is actually the pedestrian who performs abnormal behaviors. Therefore, we detect abnormal behaviors from the aspect of pedestrian rather than the aspect of scene. By adopting the aspect of pedestrian, there would be a better chance to detect more abstract abnormal behaviors.

Figure 1-1 *Our surveillance video with non-crowded and wide range scene*

In Chapter 2 of this thesis, we will first discuss some related works. In Chapter 3, we will present how we construct the module of pedestrian behavior model and how we integrate the pedestrian behavior modeling module with the detection and tracking module. After that, we will explain in details about how to detect unusual behaviors based on the proposed framework. In Chapter 4, some experimental results are presented. Finally, our conclusions are given in Chapter 5.

Chapter 2.

BACKGROUNDS

In general, most intelligent surveillance systems mainly consist of two parts. One part is tracking while the other part is abnormality detection. Tracking tracks certain regions or objects from the previous frame to the current frame. Abnormality detection, in general, makes use of tracked motion information to distinguish the salient event from daily events. Up to now, there have been plentiful techniques for tracking and abnormality detection. In this thesis, we will pay our attention mainly to those topics that are closely related to our non-crowded and complex surveillance videos.

At the beginning of this chapter, we will briefly review a few existing methods for tracking. After that, we will present some prevalent techniques for abnormality detection. Basically, most of these methods for tracking and abnormality detection ignore some important factors that may affect the movement of the tracked object. These factors include the destination of the object, the comfortable distance between objects, and the anticipated collision along the path of movement. These factors have a great impact over object's behaviors. Moreover, they may change as the environment changes. By knowing these factors, we believe we can not only track the object more accurately but also detect more complicated abnormal behaviors. Hence, at the end of this chapter, we will briefly introduce a few existing works which try to estimate these factors or to properly model object's behaviors.

2.1. Object Tracking

Object tracking has been one of the key processes in surveillance systems for a long time. Lots of researchers have put a great amount of efforts to develop techniques that can solve the correspondence problems. Establishing correspondence between frames has been thought as one of most challenging problems in object tracking, especially when occlusion occurs. Nowadays, due to the contributions of those researchers, the techniques of object tracking have evolved into much more complicated techniques. In this section, to obtain a comprehensive review of object tracking, we only introduce the basic form of some well-known tracking techniques rather than introducing the latest tracking techniques. The following discussion about object tracking is mainly based on [1].

In the following, we will introduce the basic methods for single-object tracking first. After that, we extend our discussion from single-object tracking to multiple-object tracking. We will introduce various approaches for multiple-object tracking. The main difference between single-object tracking and multiple-object tracking is the establishment of object correspondence over frames, especially when occlusion between objects occurs.

2.1.1. Single-Object Tracking

Single-object tracking is relatively simple compared to multiple-object tracking. In single-object tracking, the needs for establishing correspondence between objects are much simpler. However, up to now, there is not any tracking technique that is suitable for all cases. Each tracking technique has its pros and cons. Different tracking methods shall be adopted depending on the purpose and the circumstance. According to the approaches of these basic tracking techniques, we roughly classify them into two categories: deterministic tracking and probabilistic tracking.

2.1.1.1. Deterministic Tracking

The concept of deterministic tracking is quite intuitive. Given a region in the previous frame, we find the most similar region in the current frame. In this category, some algorithms take point-based approach, some use kernel-based approach, and others adopt silhouette approach. Below, we will introduce a few well-known tracking methods.

MEAN-SHIFT TRACKING

Mean-shift algorithm is a mathematic tool to find the local extreme value of any function. As long as an initial object appearance and a similarity metric are given, the mean-shift algorithm will maximize the appearance similarity iteratively based on the given similarity metric.

OPTICAL FLOW TRACKING

Optical flow tracking generates dense flow fields by computing the flow vector of each pixel under the brightness constancy constraint.

KLT TRACKING

As an interest point is given, KLT tracking iteratively computes the translation of given region centered on the interest point. Once the new location of the interest point is found, KLT tracking computes an affine transform to evaluate the quality of the new location of the interest point. If the quality is good, KLT tracking continuously tracking this interest point and take its nearby region as a feature. Otherwise, the interest point is eliminated.

Eigen Tracking

Eigen tracking builds a subspace representation of a given appearance by using Principal Component Analysis (PCA). Given an input frame, we reconstruct input appearance by using eigenvectors. The tracked object is found by minimizing the difference between reconstructed appearance and input appearance.

SVM Tracking

In general, SVM is a classification scheme. As a set of positive and negative training features are given, SVM finds the best hyper-plane which separates these two classes with the largest separation. In SVM tracking approach, images of tracked objects as treated as positive features while things that are not tracked are treated as negative features.

Contour Tracking

Contour tracking methods are employed when tracking a complete object is needed. Starting at an initial contour, contour tracking methods iteratively evolve an initial contour to the current contour. In order to evolve the contour correctly, contour tracking methods require some degree of object overlapping between the previous frame and the current frame.

2.1.1.2. Probabilistic Tracking

In general, observations from surveillance video contain noise. Moreover, object motions can undergo random perturbation. Hence, contrary to deterministic tracking which uses observations only, probabilistic tracking tries to model the states behind the observations. In other words, probabilistic tracking use the state-space approach to take into account the observation uncertainties as well as model uncertainties. The state-space model of probabilistic tracking is shown in Figure 2-1. The states behind observations could be object position, velocity, and acceleration.

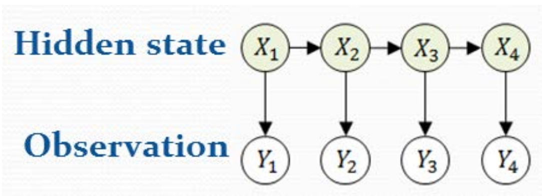

Figure 2-1 *State-space model for probabilistic tracking*

KALMAN FILTER

Kalman filter is one of the classical methods in probabilistic tracking. Kalman filter provides an efficient way to estimate the state which is governed by a linear stochastic equation and disturbed by an additive Gaussian noise. Not only the states are corrupted by an additive Gaussian noise, Kalman filter also assumes that the observations are perturbed by an additive Gaussian noise too. The assumption of additive Gaussian noises makes the closed-form solution is achievable.

Kalman filter is mainly composed of two steps: prediction step and correction step. Prediction step uses state model to predict new state, as shown in Figure 2-1. Correction step uses the current observation to update the state to decrease the problem of error propagation.

PARTICLE FILTER

However, the corrupted noises in observations and state models in reality would not be as simple as an additive Gaussian noise. Moreover, the state transition might not follow a linear dynamic equation. As a result, the closed-form solution will not be obtained easily or even be unobtainable.

Particle filter provides a solution for solving non Gaussian distribution and non-linear dynamic model by using a group of particle with different weights to approximate the distribution. Similar to Kalman filter, there are two steps in Particle filter, prediction step and correction step. Prediction step in Particle filter uses a set of particles to generate a new set of particles. Correction step in Particle filter uses observations to update the weight of each particle in new set. Since the objects are described as a probability distribution approximated by numbers of particles, we have a big chance to find the right position of object even when occlusion occurs.

2.1.2. Multiple Objects Tracking

Even though object tracking has been thought as a mature field, nowadays, a robust multiple objects tracking remains a challenging topic. One of crucial problems in multiple objects tracking is the establishment of correspondence. Unlike we discuss in single object tracking, we do not classify techniques into different categories here because most techniques we introduced in single object tracking cannot be applied directly in multiple objects tracking. Correspondence problem has to be solved first. In the following, we will briefly describe some widely used methods that deal with correspondence problem.

Joint Probability Data Association Filter (JPDAF)

The JPDAF [2], [3], is actually an extension of the Kalman filter. The JPDAF tries to associate all observations with each existing track in a probabilistic approach. Rather than choosing the nearest neighbor or data closest to what is expected, the JPDAF computes new weights for the various candidate observations. Then, the JPDAF integrates these weights into the innovation step of Kalman filter. As a result, each observation is assigned to a certain track.

Multiple Hypothesis Tracking (MHT)

Since establishing correspondence with previous frame and current frame has a great chance to be inaccurate. The core idea of MHT algorithm is to defer the correspondence decision until enough observations are measured. The MHT keeps several correspondence hypotheses at each time step and over time period, the most likely hypothesis is chosen as the final track of object. With the help of deferring the time of making correspondence decision, MHT algorithm is able to continuously track multiple objects even some observations of objects are missing.

According to these basic tracking techniques above, we can find out that most of tracking techniques do not take the interactions between objects into account. Most of these techniques only consider the difference between objects and background and solve correspondence problem or collision problem as needed.

2.2. Abnormality Behavior Detection

Automatically abnormality detection is an appealing and practical topic since it can help valuable human attention to focus on the most salient context. Many researchers have put lots of effort here. However, automatically abnormality detection remains to be a quite challenging problem. Due to its highly intellectual and abstract concepts, to give a complete definition of abnormality is even hard for us. Therefore, abnormality is usually defined in a subjective form. Sometimes, the definition of abnormality is given according to the algorithms can detect.

Since there is a great diversity of approaches to abnormal detection, in this section, we will only focus on the abnormal behaviors with non-crowded and complex scenes. In our surveillance video, most of methods to detect abnormal behaviors are based on tracked information. As we mentioned before, abnormality is an abstract concept. However, the tracked information is not able to express this concept. Namely, even though the motion trajectories can be tracked perfectly, if there is no appropriate translation, the motion trajectories would be useless. As a result, most of methods to detect unusual behaviors are based on the construction of normal scene model by training. As long as the object which contradicts this normal scene model, the object will be viewed as the occurrence of abnormal behavior. Below, we will take two examples which are similar to our surveillance videos.

[4] proposed a scene model which models the pixel-wise probability density function of object size and object velocity at training step. Each probability density function in each pixel is modeled as a multivariate Gaussian mixture function. The Gaussian mixture function is learned by using motion trajectory with EM-based learning.

As for the abnormality detection, they use two different abnormal measurements. One measurement is local abnormal analysis such as instantaneous change of velocity or size. The other one measurement is global abnormal analysis. Global abnormal analysis tries to detect abnormal behaviors which cannot be detected by local abnormal analysis such as walking onto road or glass. By using these two abnormal analysis methods, the abnormal behaviors can be detected in this surveillance system are shown in Figure 2-2. Those abnormal behaviors are walking with unusual path, riding a biking, suddenly sitting down, skating and walking on road.

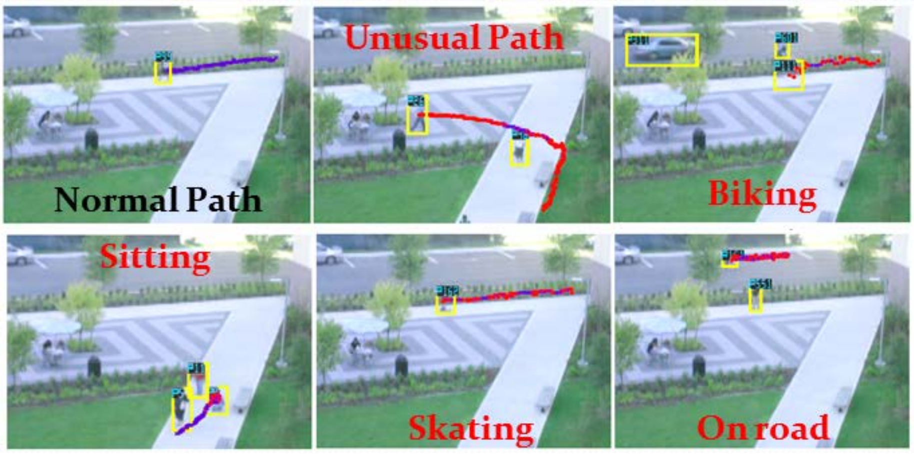

Figure 2-2 *Detected abnormal behaviors*

Another example is described in [5]. The scene model in [5] is much more complex than in [4]. [5] constructed a scene model which jointly models object appearance and dynamics of the scene. As a result, this scene model is able detect both in temporal abnormality and in spatial abnormality. They adopted the mixture of dynamic textures (MDT) to model the dynamics of scene and they used discriminant saliency to model spatial abnormality. That is, MDT is used to detect temporal abnormal behaviors which do not occur frequently and discriminant saliency is used to discriminate the appearance of object from others. The abnormal behaviors can be detected in [5] is shown in Figure 2-3. Those abnormal behaviors are driving a car, riding a bike and skating.

As we can see from those two examples of abnormal behavior detection, we can find out that the abnormal behaviors can be detected in non-crowded and wide range scenes are very similar. As most of other approaches for abnormal detection, [4] and [5] all detect abnormal behaviors from the aspect of scene. There are few approaches which detect abnormal behaviors from the aspect of pedestrian.

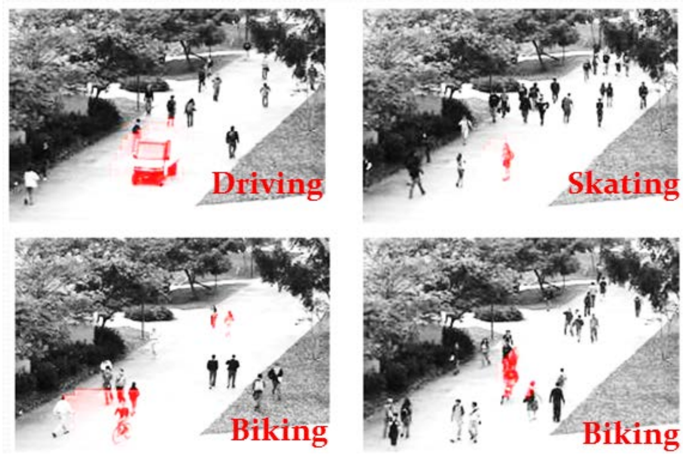

Figure 2-3 *Detected abnormal behaviors*

2.3. Pedestrian Behavior Modeling

Whenever we talk about human behavior, we might feel that human behavior is irregular and not predictable. In general, this is true for those behaviors found in complex scenes. However, we might still be able to model human behavior in some restricted and simple scene, such as in a huge population of individuals.

In order to model pedestrian behavior accurately, taking social interactions into account is an essential step. For a long time, social interactions, interactions between environment and objects and interactions between objects, have been known to affect pedestrian behavior heavily. However, these dynamic social interactions are less explored.

Modeling pedestrian behavior, especially crowd behavior, has been an important part in civil engineering and computer graphic fields and has been studied extensively in those fields. The main goal in those fields is to create realistic crowds motion in a simulated way. For example, in civil engineering [6] and [7], having pedestrian models can provide valuable information for designing and planning evacuation plans and pedestrian areas, as shown in Figure 2-4.

Figure 2-4 People in yellow shirts evacuate a building

Most of models which are used in those fields are based on the Social Force Model. The Social Force Model, proposed by Helbing and Molnar in [8] is one of the first pedestrian behavior models which take social interactions into consideration. The Social Force Model tries to capture the effects of neighboring pedestrians and scene on each individual pedestrian. The Social Force Model describes the effects, or so-called social factors, by using a combination of different social forces. These different social forces that drive pedestrians toward their goals are analogous to the real forces which apply on moving object in reality. This model allows a large scale view of large crowds and has been extended to describe various kinds of crowds [9]. In [9] proposed a combination of social panic model and social force model to create a more generalized model.

Modeling pedestrian behavior in computer vision has received much more attentions in past few years since we can get lots of advantages from an accurate pedestrian behavior model. Having an accurate pedestrian model not only improve the tracking performance but also improve the performance of abnormal behavior detection.

Rather than using a simulated way to model pedestrian behavior, most of models in computer vision use real-world data to produce an accurate pedestrian behavior model. One of earliest pedestrian behavior models to take advantage of behavioral priors in tracking problems is Discrete Choice Model in [10].

2.3.1. Discrete Choice Model (DCM)

Discrete Choice Model, proposed by Antonini et al. in [10], uses a series of specific discrete choices to model pedestrian behavior. The next position of individual pedestrian is decided as a sequence of discrete choices is made. At each time instant, they dynamically discretize spatial locations for each individual pedestrian. The most probable location in next time step for individual pedestrian is chosen based on social factors in the scene from a discrete choice set, as shown in Figure 2-5. The DCM uses 33 possible locations in the discrete choice set. That is, the next position of pedestrian locates on one of thirty-three discrete positions.

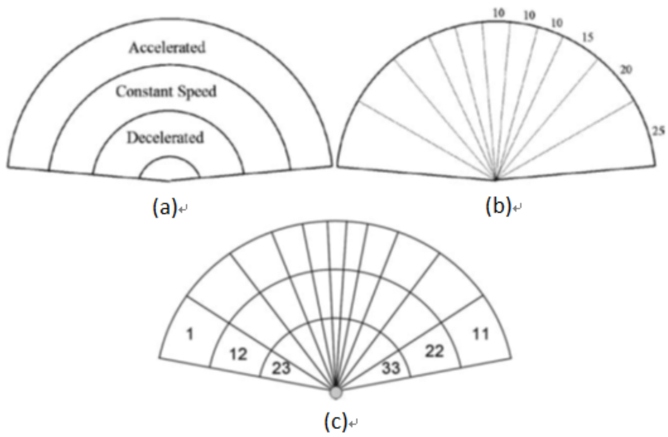

Figure 2-5 Choice Set. (a) Speed choice set. (b) Orientation choice set. (c) Choice set of combining (a) and (b).

However, discretization in possible positions introduces some difficulties. One of the difficulties is size of discretization grid. Fine spatial discretization increases not only the accurate prediction but also the complexity of computation substantially. However, coarse spatial discretization affects the accuracy of prediction. Making a balance between fine grid and coarse grid is quite challenging.

Contrary to discrete grid of spatial position, Linear Trajectory Model has been proposed recently in continuous form in [11].

2.3.2. Linear Trajectory Avoidance (LTA)

Linear Trajectory Avoidance (LTA) is a model which takes scene information and interactions between different targets into account. LTA model considers three social factors which influence pedestrian behavior heavily. These factors are all objects in scene, constant velocity and the destination. Based on these three factors, LTA constructs the overall energy for each pedestrian. The next time position of pedestrian is predicted by minimizing the overall energy as white dots in Figure 2-6.

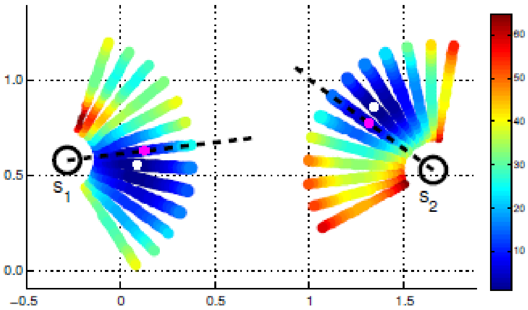

Figure 2-6 *Illustration for the predicted position for each pedestrian. Magenta points are the predicted position without taking social interactions into account. White points are the predicted position with taking social interactions into account.*

Figure 2-6 shows that if the pedestrians take social interactions into account, the phenomenon of avoidance will happen as shown in white dots in Figure 2-6. That is, the left-hand side pedestrian will slow down and turn right slightly. However, the right-hand side pedestrian will accelerate and turn right slightly.

The LTA model has been shown to work well for multiple objects tracking in complex scenarios. However, there are some limitations of the current LTA model. An example of limitations shown in [12] is the pedestrian is walking as part of a group, as shown in Figure 2-7. The LTA model does not model groups of pedestrians since the behavior of pedestrian in a group is quite different from the behavior of single pedestrian.

Figure 2-7 *An example of limitations of the LTA model.*

Estimating social relationship has received much more attentions recently [13] and [14]. Most of them only try to explore the group relation of a crowd in photos or in surveillance videos. However, there are still few methods to explore social group and put it into tracking scheme.

Chapter 3.

PROPOSED METHOD

Most research works in abnormality detection focus on the construction of scene model. The scene model could be clusters of normal trajectories or distributions of normal object size and normal object velocity. All these approaches try to obtain the scene knowledge with normal behaviors or events to detect abnormalities.

In our thesis, instead of obtaining the scene knowledge, we detect abnormalities from the aspect of pedestrian. We believe that more abstract abnormalities can be detected from this aspect. Figure 3-1 shows the flow chart of our abnormal surveillance system.

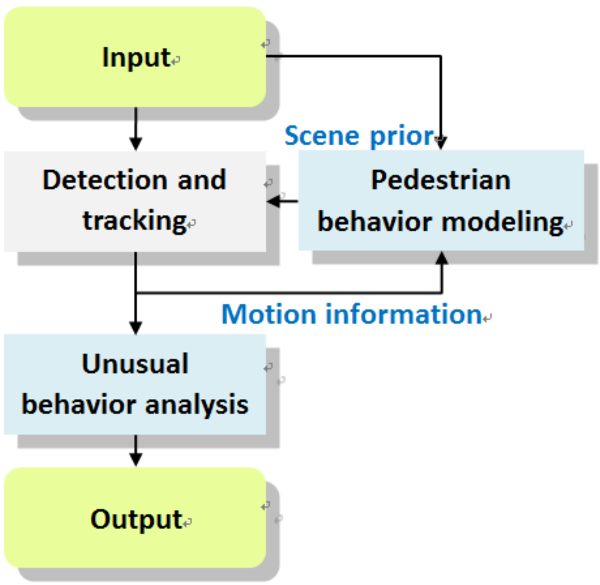

Figure 3-1 *Flow chart of our proposed surveillance system*

In our system, we try to integrate a module of pedestrian behavior modeling into the original module of detection and tracking. This module of pedestrian behavior modeling describes the social interactions between scene and pedestrians and the social interactions between pedestrians. Finally, we detect unusual behaviors based on this introduced module of pedestrian behavior modeling. At the beginning of our system, we make an assumption to simplify problems. We assume that the pedestrians who appear at each time step can all be detected with an additive Gaussian noise perturbation. That is, we do not try to solve occlusion and correspondence problems in our system. As a result, in this thesis, we will mainly focus on how to construct our module of pedestrian behavior modeling and how to detect unusual behaviors based on this module of pedestrian behavior modeling.

We organize this chapter as follows. At the beginning of this chapter, we will present our purposed graphical representation which shows the approach we take the module of pedestrian behavior modeling into pedestrian tracking. Then, we will discuss how we construct the module of pedestrian behavior modeling by estimating pedestrian properties. After that, we will talk about how to combine pedestrian properties with detection and tracking module. Finally, we will discuss how to make use of this module of pedestrian behavior modeling to detect unusual behaviors and even to detect abnormal behaviors.

3.1. Graphical Representation

As mentioned before, we detect abnormal behaviors mainly based on the module of pedestrian behavior modeling and this module will describe the social interactions between surrounding and objects social interactions between objects. However, modeling pedestrian behavior directly is difficult since there are lots of pedestrian behaviors. As a result, we explore the social factors which have a great impact on pedestrian behavior first instead. The social factors we used here are personal properties.

This approach can be well-understood by using a graphical representation. In order to take the effect of social factors, which lead to various behaviors of pedestrian, into account, we introduce an additional state layer above the original graphical model of Kalman filter in our tracking module, as shown in Figure 3-2. Our proposed graphical model actually is quite similar to the switching Kalman filter as shown in Figure 3-2. In fact, the functionality of our proposed additional state layer is analogous to the switching state in the switching Kalman filter. The introduced additional state layer in our proposed structure is used to describe various hidden properties of pedestrian which lead to lots of different pedestrian behaviors and switching state in the switching Kalman filter tries to model different possible causes.

These hidden properties of pedestrian in our proposed structure are actually the thoughts of pedestrian, such as comfortable distance, preferred walking velocity and destination etc. These properties are much more abstract than usual tracked information, like exact position or exact velocity and effect pedestrian behavior heavily. With the help of these hidden properties of pedestrian which already consider the social factors, we are able to accurately model the behavior of pedestrian.

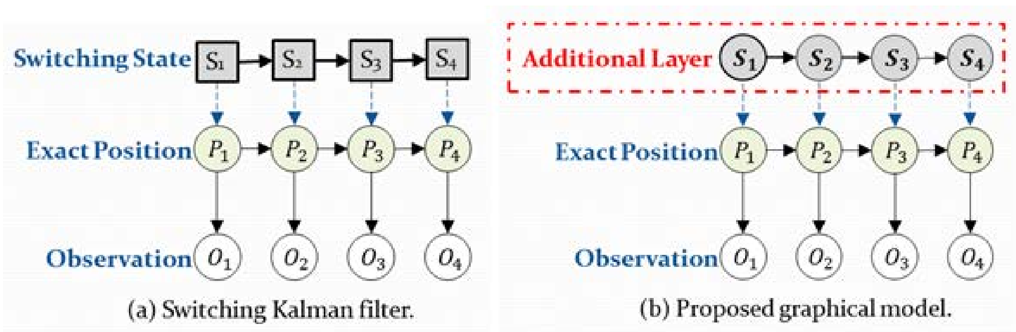

Figure 3-2 *(a) Graphical model of switching Kalman filter (b) Graphical model of proposed method*

Even though the graphical model of our proposed graphical model and the switching Kalman filter looks like the same, there are two main differences between these two graphical models. First, the switching state in traditional switching Kalman filter is usually a discrete random variable. This additional state layer in our structure however, is a set of deterministic variables and random variables. Second, as we are handling multiple pedestrians, these states in our additional state layer of our structure are no longer independent. There are some interactions between these additional states as shown in Figure 3-3.

These structures, our proposed structure and the switching Kalman filter, in general have greater descriptive power than the traditional Kalman filter since it contains multiple choices. However, having an exact inference is intractable not only in our proposed structure but also in the switching Kalman filter. We have to make some assumptions to simplify our proposed structure to make inference properly.

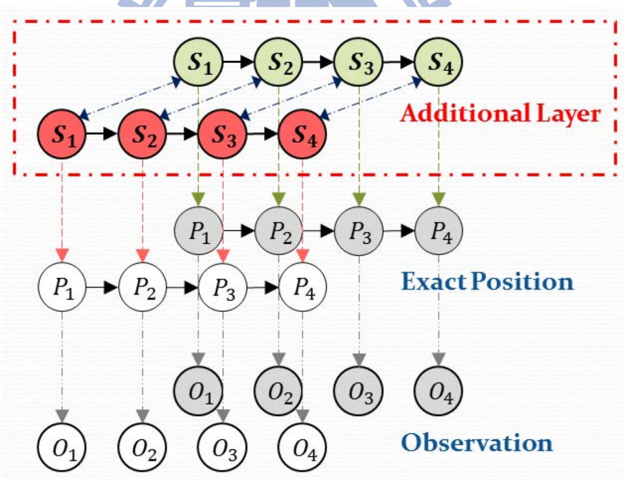

Figure 3-3 *Proposed graphical model for our multiple pedestrians*

The simplified method we use is to decouple our proposed structure into two independent and small structures as shown in Figure 3-4 and both of these two structures are computation feasible. One structure is the additional state layer portion or switching portion which contains the property of switching, and the other structure is the original Kalman filter portion which includes linear and dynamic properties of the system.

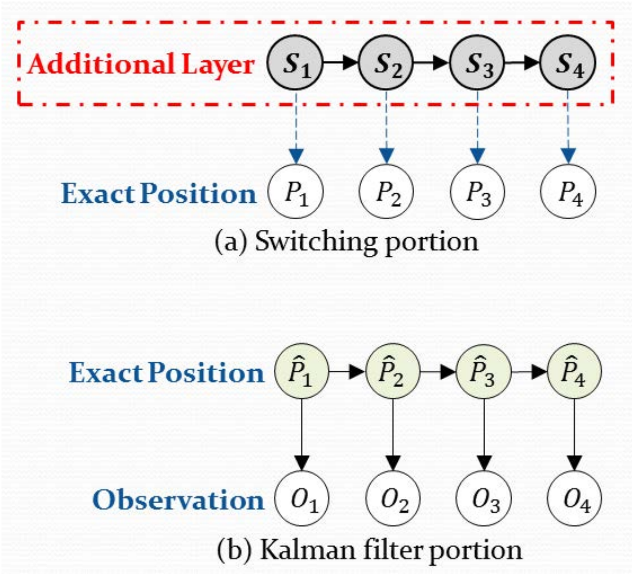

Figure 3-4 *Decouple our proposed graphical model (a) Switching portion. (b) Kalman filter portion.*

This decoupled technique, in fact, is just a variational inference technique for Bayesian network [15]. Namely, we construct a η-parameterized distribution which is close to the exact distribution but is computation feasible.

In general, the variables in additional state layer are actually all random variables. The reason why we describe the additional state layer in our proposed as a set of deterministic variables and random variable is we are not handle large random variables at the same time. If we want to estimate these large random variables at the same time, the uncertainty of each random variable will increase. This result is not what we want. Therefore, we fix some random variables and we can also view these fixed random variables as deterministic variables. The value of these deterministic variables will be trained beforehand in training step.

With the help of these simplified methods, we are able to make inference under our proposed structure. The first step of our proposed structure is to estimate the rest random variables. After obtaining these random variables as well as deterministic variables, we transform these variables into position information of pedestrian by using a set of energy functions in switching portion, as shown in Figure 3-4. Then we pass the estimated position information of pedestrian from the switching portion to the Kalman filter portion, as shown in Figure 3-4. We will make use of the estimated position information in Kalman filter portion to help us detect and track pedestrian. These two portions correspond to the two following sections of pedestrian behavior modeling and pedestrian tracking, respectively.

In following sections, we will discuss how to estimate these random variables in switching portion first. Then we will talk about the tracking scheme in Kalman filter portion.

3.2. Pedestrian Behavior Modeling

As we mentioned in previous section, we introduce an additional state layer to describe the hidden properties of pedestrian. These hidden pedestrian properties, such as comfortable distance and destination, have been known to have a great impact on the behavior of pedestrian. Since there are varieties of pedestrian behaviors, modeling pedestrian behavior directly is quite difficult. Instead of modeling pedestrian behavior directly, we try to estimate the hidden pedestrian properties first. With the help of hidden pedestrian properties, we have better knowledge about the pedestrian and what kind of behavior is going to be performed. Then, we begin to further model pedestrian behavior.

Our proposed method to model pedestrian behavior is nothing but a linear combination of a set of energy functions. These energy functions are parameterized by the estimated hidden personal properties. By using these energy functions, we turn these hidden personal properties into an analytical and tractable form. An energy map is constructed by combining different energy functions. By minimizing the energy map, we are able to predict the next time step position of the pedestrian. That is, we are capable of modeling pedestrian behavior as we make use of this energy map well.

The discussions in this section focus on the inference of decoupled switching portion as shown in Figure 3-4.

3.2.1. Personal Property Estimation

Personal properties are known to have a great influence on personal behavior and are viewed as random variables in our proposed structure. To avoid estimating large random variables at the same time, we fix some of variables in our structure. Namely, the fixed variables are deterministic variables. As for those remaining random variables, we estimate them by using data-driven method. Below, we will talk about which variables are fixed while which are not. We will also talk about how to determine or estimate these variables.

SCENE KNOWLEDGE

The behavior of pedestrian, in general, obeys some rules which are established by society. For example, normally we do not climb trees or street lights and we also do not enter the region where prohibitory enter sign is set up. Having scene prior knowledge is useful to detect those behaviors which do not occur often.

The scene knowledge we adopt in our proposed system is locations of obstacles and restricted region. We assume that the scene knowledge is known in advance and is labeled manually.

DESTINATION

Pedestrian always moves toward his destination if there is no accidental event happened. Knowing the pedestrian destination is useful for us to predict his next move. The importance of pedestrian's destination is shown in [11]. [11] also shows that even roughly guess the destination helps to make more accurate prediction.

The possible destinations, in fact, can be trained at the beginning of surveillance system by simply recording the exit and entry points of scene. However, we do not take destination information into our proposed structure. We treat it as the optional information. Because our scene is relatively simple and most abnormal events we used are designed, the destination information in our proposed system is not as useful as in daily life scene.

COMFORTABLE DISTANCE

Comfortable distance is used to describe the avoidance phenomenon as pedestrian confront other unfamiliar objects. Everyone feels different comfortable distance. Knowing the comfortable distance of pedestrian improves the prediction of pedestrian behavior since next possible actions which pedestrian might take are under our controlled.

However, estimating comfortable distance for everyone seems to be complex and difficult not only because the comfortable distance for everyone is different but also because the comfortable distance might vary as the surrounding changes. As a result, we make a simplified assumption here that everyone has the same comfortable distance and we take it as a deterministic variable. The exact value of this comfortable distance is obtained in training step.

PREFERRED VELOCITY

The property of preferred velocity is very similar to the property of comfortable distance. That is, everyone has his own preferred walking velocity and it varies as surrounding changes. Estimating preferred walking velocity for each one sounds impracticable which is just like the comfortable distance.

Therefore, we take previous observed velocity as the preferred velocity of this object since we believe that the object always tries to maintain a constant velocity. To be more specific, if the object is individual pedestrian, we take his previous velocity as his preferred velocity; if the object is a group consists of several pedestrians, we take group's previous mean velocity as the group's preferred velocity.

SOCIAL GROUP

Knowing the social group is important to model pedestrian behavior, since the behavior of pedestrian in groups tends to be very different from the behavior of pedestrian in single. In general, pedestrian behavior in group is hard to predict, however, the whole group behavior, in most case, is relatively simple and easy to model. By tracking object in group unit, tracking performance can be improved drastically.

However, estimating the social group is not an easy job because of often changing of group size. Therefore, we use pair-wise feature with bottom-up grouping to solve the problem of various group sizes. That is, we take features from every two pedestrians. Then, we adopt Support Vector Machine (SVM) to help us classify whether these two pedestrians belong to the same group. We extract eight features which are illustrated in Figure 3-5 and summary at Table 3-1. These features are the time difference of showing up, relative distance, velocity magnitude differences of past two steps, velocity orientation differences of past two steps and number of neighboring pedestrians.

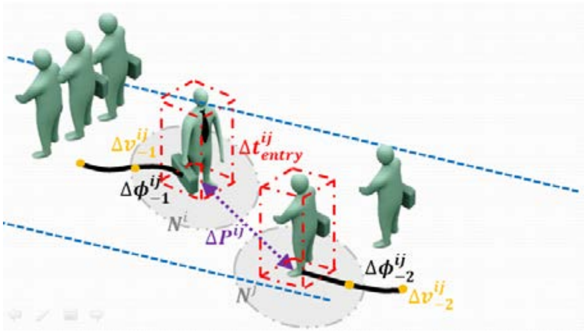

Figure 3-5 *Illustration of extracting group features*

Table 3-1 *Features for group estimation*

No.	Notation	Feature Description
1	$\Delta \boldsymbol{t}_{entry}^{ij}$	Normalized time difference of showing up
2	$\Delta \boldsymbol{P}^{ij}$	Relative distance
3,4	$\Delta \boldsymbol{v}_{-1}^{ij}$, $\Delta \boldsymbol{v}_{-2}^{ij}$	Normalized past two magnitudes of speed
5,6	$\Delta \phi_{-1}^{ij}$, $\Delta \phi_{-2}^{ij}$	Past two directions of speed
7,8	$\boldsymbol{N}^{i}$, $\boldsymbol{N}^{j}$	The number of nearby objects

The output of SVM is a binary number. In other words, it shows that whether these two pedestrians are in the same group. Based on SVM result, we group pedestrians from bottom to top. Since we believe that connected pedestrians are always exist to connect individual pedestrian to a large group as shown in Figure 3-6. At the end of grouping, we will obtain the best group size which is mainly determined by SVM.

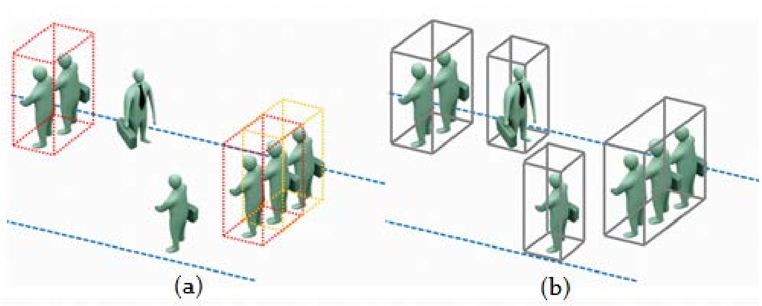

Figure 3-6 *Illustration of bottom-up grouping (a) Pair-wise grouping (b) Final group size*

3.2.2. Energy Functions for Tracking

Suppose we are able to obtain those hidden pedestrian properties by estimation or by training. The next step is to transform these abstract properties into tractable and concrete descriptions. The method we adopt is the same as in [11], which uses a set of energy functions. Different properties correspond to different energy functions. Then, we construct an energy map by simply combining different energy functions in linear way in Eq. 3-1. These energy functions we used are Avoidance Energy ($\boldsymbol{E}_{av}$)、Constant Speed Energy ($\boldsymbol{E}_{cv}$)、Scene Prior Energy ($\boldsymbol{E}_{scan}$) and optional Destination Energy ($\boldsymbol{E}_{dest}$).

$$\boldsymbol{E}_{total}^{i} = \boldsymbol{w}_{av}\boldsymbol{E}_{av}^{i} + \boldsymbol{w}_{cv}\boldsymbol{E}_{cv}^{i} + \boldsymbol{w}_{scen}\boldsymbol{E}_{scen}^{i} + \left(\boldsymbol{w}_{dest}\boldsymbol{E}_{dest}^{i}\right) \qquad \text{Eq. 3-1}$$

We have to emphasis that our tracking unit is each group rather than each individual pedestrian. That is, the index $\boldsymbol{i}$ in Eq. 3-1 represents each group rather than represents each pedestrian. The Eq. 3-1 can also be viewed as an objective function for each pedestrian. Most of methods only use the information of constant velocity which is Constant Speed Energy ($\boldsymbol{E}_{cv}$). Now, we add additional constraints on this pedestrian which are Avoidance Energy ($\boldsymbol{E}_{av}$) and Scene Prior Energy ($\boldsymbol{E}_{scan}$). These two additional terms can be taken as regularization terms. In this remaining section, we will detail the mathematical expressions of different energy functions.

3.2.2.1. Avoidance Energy

Pedestrians will always take actions in advance to avoid the collision with other objects and we use the Avoidance Energy to describe this phenomenon. The Avoidance Energy for a specific pedestrian in the scene is composed of other single pedestrian and other pedestrians in group as expressed in Eq. 3-2.

$$\boldsymbol{E}_{av}^{i} = \sum_{i \neq j,\ \ j \in S} \boldsymbol{E}_{single}^{j} + \sum_{i \neq j,\ \ j \in G} \boldsymbol{E}_{group}^{j} \qquad \text{Eq. 3-2}$$

Single Avoidance Energy

For each pedestrian, we use a Gaussian function as shown in Figure 3-7 as our energy function.

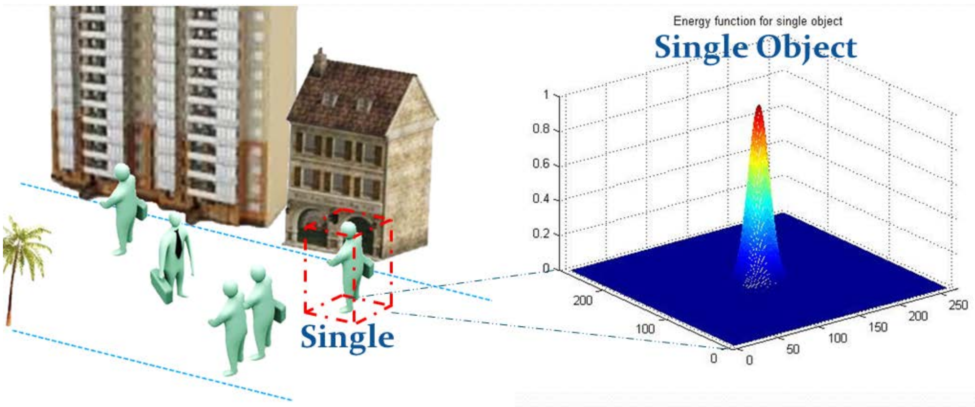

Figure 3-7 *Energy function of single pedestrian*

The mathematical expression is in Eq. 3-3 and σ_s represents the comfortable distance of pedestrian in Eq. 3-3. This Gaussian function is centered on the predicted location at next time step in Eq. 3-4 and $\boldsymbol{P}^i$ represents the location of pedestrian and $\tilde{v}^i$ represents the predicted velocity of pedestrian in Eq. 3-4.

$$\boldsymbol{E}_{single}^{i} = exp\left(-\frac{\boldsymbol{d}^{i^2}}{2\boldsymbol{\sigma}_s^2}\right) \qquad \text{Eq. 3-3}$$

$$\boldsymbol{d}^{i^2} = \left\|\boldsymbol{X} - \left(\boldsymbol{P}^i + \tilde{\boldsymbol{v}}^i\right)\right\|^2 \qquad \text{Eq. 3-4}$$

Group Avoidance Energy

As for group object, we take mixture of Gaussian function as our energy function as shown in Figure 3-8.

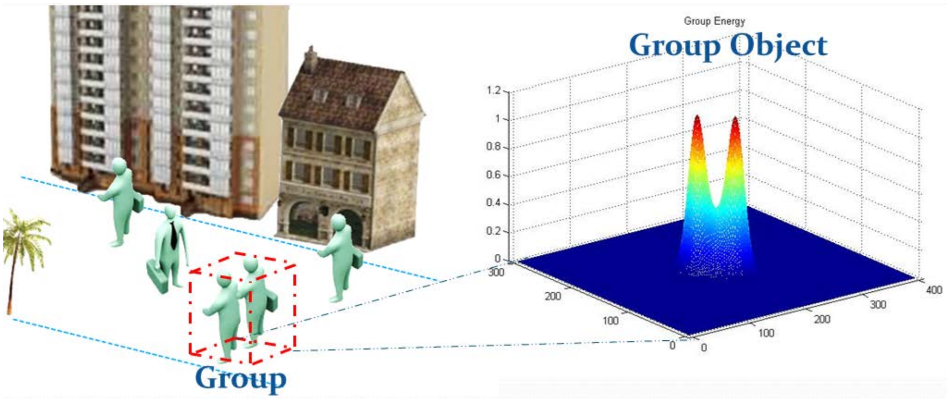

Figure 3-8 *Energy function of group*

This mixture of Gaussian function is simply the Gaussian function for each pedestrian plus an additional weighted Gaussian function in Eq. 3-5. This mixture of Gaussian function is centered on predicted group center at next time step in Eq. 3-6 and Eq. 3-7.

$$\boldsymbol{E}_{group}^{i} = exp\left(-\frac{\boldsymbol{d}^{i^{2}}}{2\boldsymbol{\sigma}_{s}^{2}}\right) + exp\left(-\frac{\boldsymbol{d}^{j^{2}}}{2\boldsymbol{\sigma}_{s}^{2}}\right) + \boldsymbol{w}_{g}\, exp\left(-\frac{\boldsymbol{d}^{ij^{2}}}{2\boldsymbol{\sigma}_{g}^{2}}\right) \qquad \text{Eq. 3-5}$$

$$\boldsymbol{d}^{k^{2}} = \|\boldsymbol{X} - (\boldsymbol{P}^{k} + \tilde{\boldsymbol{v}}^{G})\|^{2}, \quad k = i, j \qquad \text{Eq. 3-6}$$

$$\boldsymbol{d}^{ij^{2}} = \left\|\boldsymbol{X} - \left(\boldsymbol{P}^{ij} + \tilde{\boldsymbol{v}}^{G}\right)\right\|^{2} \qquad \text{Eq. 3-7}$$

3.2.2.2. Constant Velocity Energy

Pedestrian in single or in group, in general, always move in a constant velocity. The Constant Velocity Energy is used to constrain the object to maintain a constant velocity. Our Constant Velocity Energy mainly consists of two components, Magnitude Energy and Orientation Energy. We combine these two components in a linear way as presented in Eq. 3-8.

$$\boldsymbol{E}_{cv}^{i} = \boldsymbol{E}_{mag}^{i} + \boldsymbol{w}_{ori}\, \boldsymbol{E}_{ori}^{i} \qquad \text{Eq. 3-8}$$

The Constant Velocity Energy for each object is shown in Figure 3-9. The color of map shows that the object's energy at each location. The lower energy, the more probable of the pedestrian shows up at next time step. That is, the pedestrian always moves toward to the position where the energy is the lowest.

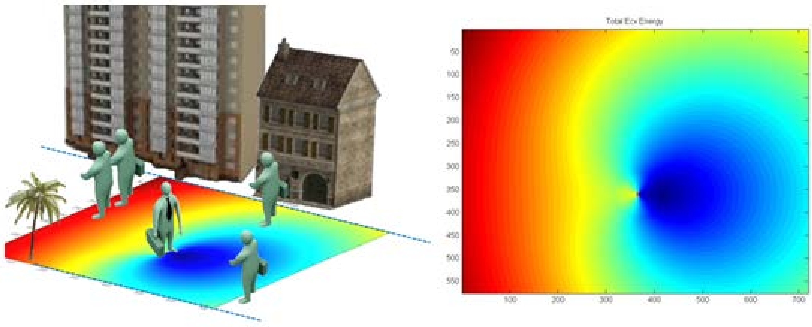

Figure 3-9 *Constant Velocity Energy*

Since the property of constant velocity belongs to the pedestrian who is being tracked, this energy function of pedestrian in group will be slightly different than the energy function in single. Therefore, we will talk about both cases in detail below.

MAGNITUDE ENERGY

Since object prefers to keep same magnitude as in previous step, the Magnitude Energy is calculated by computing the deviation of velocity magnitude from the magnitude of preferred velocity and taking absolute value of deviation as presented in Eq. 3-9 or Eq. 3-10. Eq. 3-9 and Eq. 3-10 represent single object and group object of the Magnitude Energy respectively. In Eq. 3-9, u^{i} is the preferred velocity of single pedestrian and u^{G} in Eq. 3-10 is the preferred velocity of group.

$$\boldsymbol{E}_{mag}^{i} = abs(\|\boldsymbol{v}\| - \|\boldsymbol{u}^{i}\|), \qquad \text{Eq. 3-9}$$

$$\boldsymbol{E}_{mag}^{i} = abs(\|\boldsymbol{v}\| - \|\boldsymbol{u}^{G}\|), \qquad \text{Eq. 3-10}$$

Figure 3-10 shows the figure of Magnitude Energy. It makes sense that the figure of Magnitude Energy is non-oriented and sensitive to magnitude.

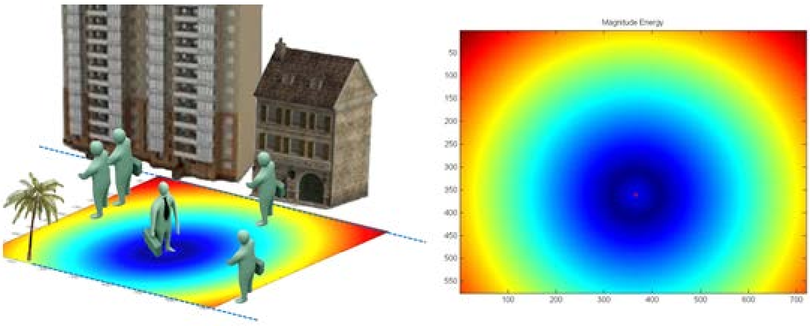

***Figure 3-10** Figure of Magnitude Energy*

<u>ORIENTATION ENERGY</u>

Since object prefers to move in same direction as in previous step, we calculate the deviation of angle from the direction of preferred velocity first as expressed in Eq. 3-11 or Eq. 3-12. Eq. 3-11 and Eq. 3-12 represent single object and group object respectively. In Eq. 3-11, P^i is the current location of single pedestrian and P^G in Eq. 3-12 is the center location of group.

$$\phi = \frac{(\boldsymbol{X} - \boldsymbol{P}^i) \cdot \boldsymbol{u}^i}{\|\boldsymbol{X} - \boldsymbol{P}^i\| \times \|\boldsymbol{u}^i\|} \qquad \text{Eq. 3-11}$$

$$\phi = \frac{(\boldsymbol{X} - \boldsymbol{P}^G) \cdot \boldsymbol{u}^G}{\|\boldsymbol{X} - \boldsymbol{P}^G\| \times \|\boldsymbol{u}^G\|} \qquad \text{Eq. 3-12}$$

Then, we use a level-shifted cosine function as our energy function for the orientation energy in Eq. 3-13.

$$\boldsymbol{E}^i_{ori} = \frac{\mathbf{1} + cos(\phi)}{\mathbf{2}} \qquad \text{Eq. 3-13}$$

Figure 3-11 shows the figure of Orientation Energy. It is reasonable that the figure of Orientation Energy is oriented and non-sensitive to magnitude.

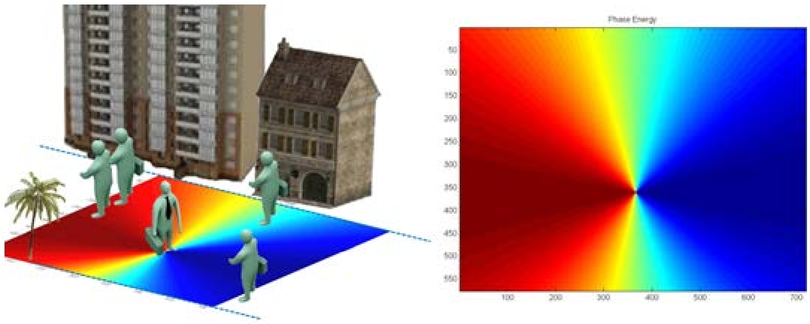

Figure 3-11 *Figure of Orientation Energy*

3.2.2.3. Scene Prior Energy

As mentioned before, the scene knowledge we used is the locations of obstacles and the restricted regions. We model energy function of obstacles simply by using a Gaussian function and we use a level-shift constant as an energy function for the restricted regions which are illustrated respectively in Figure 3-12.

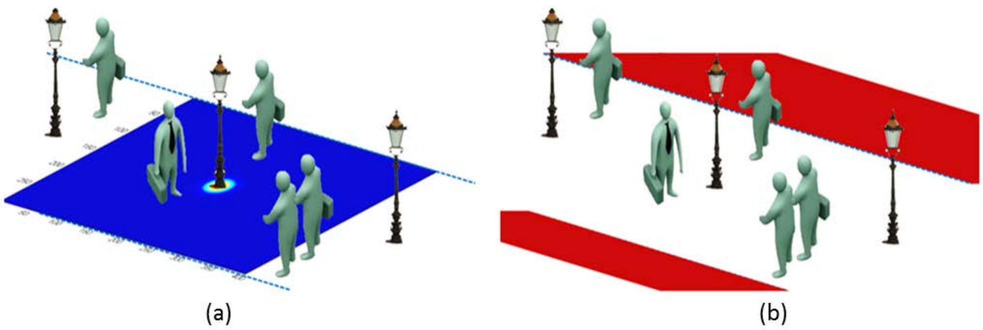

Figure 3-12 *Illustration of scene prior energy (a) Obstacle energy. (b) Restricted region energy*

3.2.3. Parameters Optimization

Finally, there is still one thing we left behind without discussion. That is, how to decide the values of deterministic variables. As we talked before, every hidden property actually can be viewed as a random variable. Since we are not able to handle such large unknowns at the same time, we fix some random variables to be deterministic values. These variables, deterministic variables, are comfortable distance of individual (σ_s) and comfortable distance of groups (σ_g).

These deterministic variables as well as those weighted coefficients, w_{ori}, w_g and w_{cv}, are determined by training data in training step. In training step, we take normal surveillance videos and construct an energy map for each pedestrian respectively. We estimated each pedestrian position by finding the minimum location of the energy map in Eq. 3-14. Then, we compute least-square error by comparing the estimated position with ground truth observation in Eq. 3-15. We choose those parameters with smallest least square error.

$$\widehat{\boldsymbol{P}}^{i}_{n|n-1} = \arg\min_{\boldsymbol{P}} \boldsymbol{E}^{i}_{total}(\boldsymbol{\Theta}) \qquad \text{Eq. 3-14}$$

$$\arg\min_{\boldsymbol{\Theta}} \left\| \boldsymbol{O} - \widehat{\boldsymbol{P}}^{i}_{n|n-1}(\boldsymbol{\Theta}) \right\|^{2} \qquad \text{Eq. 3-15}$$

3.3. Pedestrian Tracking

As we mentioned before, we decouple our proposed structure into switching portion and Kalman filter portion. In this section, the following discussions will focus on the inference of decoupled Kalman filter portion as shown in Figure 3-4.

Although the graphical model looks like the same as Kalman filter as shown in Figure 3-13, there is one main difference between these two models. In the decoupled Kalman filter portion, there are some parameters that are passed down from switching portion but traditional Kalman filter does not have these parameters as shown in Figure 3-13.

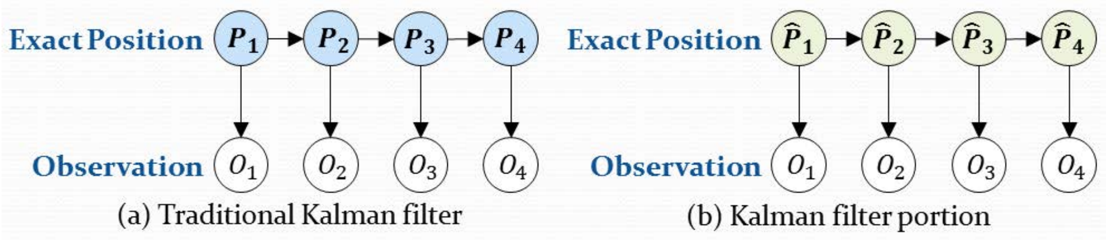

Figure 3-13 *Graphical representation (a) Traditional Kalman filter (b) Our Kalman filter portion*

The parameter comes from switching portion in our proposed structure is the position information of pedestrian. Following, we will briefly introduce the basic operations of traditional Kalman filter first. Then we will turn our attentions to our decoupled Kalman filter portion.

As we mentioned in Chapter 2, traditional Kalman filter is a very classical statistic tracking technique. As shown in Figure 3-13, traditional Kalman filter has two state layers. One state layer is the hidden state, which describes the hidden information behind the observations, and the other state layer is the observation. The relationship between these states can be completely characterized by two linear equations with a Gaussian additive noise respectively in Eq. 3-16 and Eq. 3-17.

$$\mathbf{P}_n = \boldsymbol{A}\boldsymbol{P}_{n-1} + \boldsymbol{u}_n \qquad \text{Eq. 3-16}$$

$$\boldsymbol{O}_n = \boldsymbol{H}\boldsymbol{P}_n + \boldsymbol{w}_n \qquad \text{Eq. 3-17}$$

By predicting and estimating the evolution of hidden state, traditional Kalman filter is able to detect objects in Eq. 3-18 and Eq. 3-19.

$$\hat{\mathbf{P}}_{n|n-1} = \boldsymbol{A}\hat{\boldsymbol{P}}_{n-1|n-1} \qquad \text{Eq. 3-18}$$

$$\hat{\boldsymbol{O}}_n = \boldsymbol{H}\hat{\boldsymbol{P}}_{n|n-1} \qquad \text{Eq. 3-19}$$

However, since the predicted state $\hat{P}_{\mathrm{n|n-1}}$ in our decoupled Kalman filter comes from the switching portion, the predicted equation is a little different from the original one in Eq. 3-18. The change of $\hat{P}_{\mathrm{n|n-1}}$ leads to a time-vary transition matrix $\hat{A}_{\mathrm{n}}$ in Eq. 3-20.

$$\hat{\boldsymbol{P}}_{n|n-1} = \hat{\boldsymbol{A}}_n \hat{\boldsymbol{P}}_{n-1|n-1} \qquad \text{Eq. 3-20}$$

Except for the time-vary transition matrix, the rest equations in our decoupled Kalman filter are exactly the same as in traditional Kalman filter. For example, the correction of predicted state is described in Eq. 3-21, the Kalman coefficient is expressed in Eq. 3-22 and MSE matrix is shown in Eq. 3-23 and Eq. 3-24 are the same as in traditional Kalman filter.

$$\hat{\boldsymbol{P}}_{n|n} = \hat{\boldsymbol{P}}_{n|n-1} + \mathbf{K}_n\left(\boldsymbol{O}_n - \boldsymbol{H}\hat{\boldsymbol{P}}_{n|n-1}\right) \qquad \text{Eq. 3-21}$$

$$\mathbf{K}_n = \mathbf{M}_{n|n-1}\boldsymbol{H}^{\mathrm{T}}\left(\boldsymbol{C}_n + \boldsymbol{H}\mathbf{M}_{n|n-1}\boldsymbol{H}^{\mathrm{T}}\right)^{-1} \qquad \text{Eq. 3-22}$$

$$\mathbf{M}_{n|n-1} = \hat{\boldsymbol{A}}_n \mathbf{M}_{n-1|n-1} \hat{\boldsymbol{A}}_n^{\mathrm{T}} + \mathbf{Q}_{\mathrm{u}} \qquad \text{Eq. 3-23}$$

$$\mathbf{M}_{n|n} = (\mathbf{I} - \mathbf{K}_n \boldsymbol{H})\mathbf{M}_{n|n-1} \qquad \text{Eq. 3-24}$$

3.4. Unusual Behavior Detection

Until now, we are able to track pedestrians with their personal properties. The next step is to detect unusual behaviors based on their personal properties.

The approach we use to detect unusual behaviors is to calculate the difference between two estimated personal properties and observation. To more specific, we detect unusual behaviors by comparing difference between two maps. These maps are constructed by using the Avoidance Energy function which is centered at each pedestrian position. That is, at each time step, from those estimated personal properties we are able to predict the next time position of each pedestrian. Then we construct a map based on the position of each pedestrian. At the same time, we observe the position of detected pedestrians and we are able to produce another map based on the observed position. The difference between these two maps reveals the behavior to which we should pay more attention as shown in Figure 3-14.

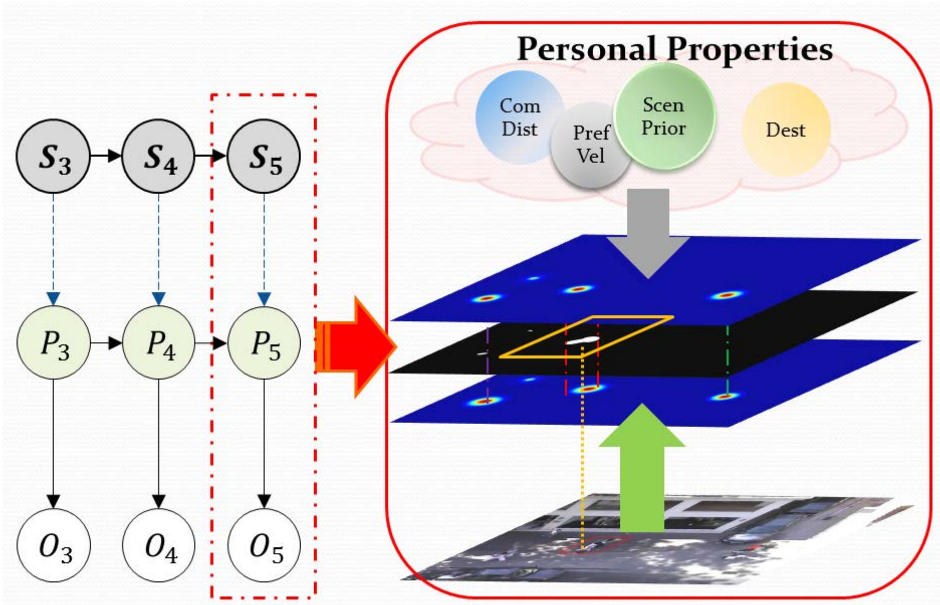

Figure 3-14 *Illustration of finding unusual behaviors*

The intuitive perspective of computing difference of two maps is that we hope that all the social factors are included in our estimated personal properties and the behavior of pedestrian should follow our estimated behavior. However, the exact observed behavior of pedestrian is quite different from our estimated behavior. This means that we should beware of this observed behavior since some unknown factors which we do not take into consideration at current time step affect the pedestrian. This observed behavior which differs from predicted behavior is called unusual behavior in our thesis.

However, the detected unusual behavior has an extreme high probability to be a normal behavior. For example, the pedestrian might suddenly get a call and slow down his paces in order to response the call. Since our estimated personal properties are not able handle such kinds of accidental events, this behavior will be detected as an unusual behavior by our proposed method. However, this behavior of pedestrian is pretty normal and trivial. Therefore, it is necessary for us to further analysis what kind of behavior is preforming.

In order to understand what behavior is performing, we take a spatiotemporal patch as a feature to help us distinguish abnormal behaviors from those normal, daily behaviors, as illustrated in Figure 3-15.

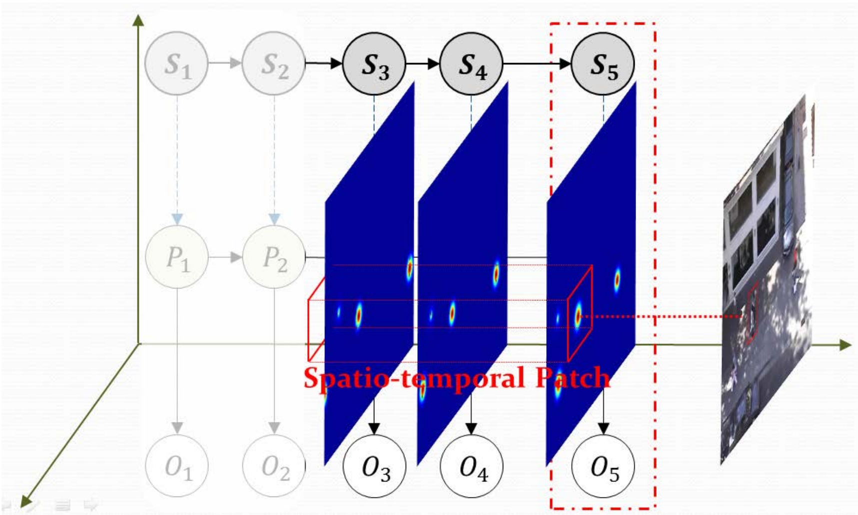

Figure 3-15 *Illustration of extracting abnormal features*

As long as the unusual behavior is detected, we will recognize who performs this unusual behavior. Then, we take a spatiotemporal patch which centers on the unusual pedestrian from the last two energy maps and current energy map. However, the size of this spatiotemporal patch varies which depends on the past motion of the unusual pedestrian. To make use of this spatiotemporal feature, we have to normalize this feature to the same standard. The normalization step is illustrated in Figure 3-16. First, we convert this spatiotemporal patch to a binary patch and use sign distance transformation. Then we normalize this patch to a 20-by-40 size patch. This normalized patch will be taken as an input of Support Vector Regression (SVR). SVR compares this normalized feature to those pre-defined, well-labeled features and gives out a predicted value. This predicted value indicates which pre-defined behavior is the most probable. The pre-defined and well-labeled features are extracted in training step as shown in Figure 3-17.

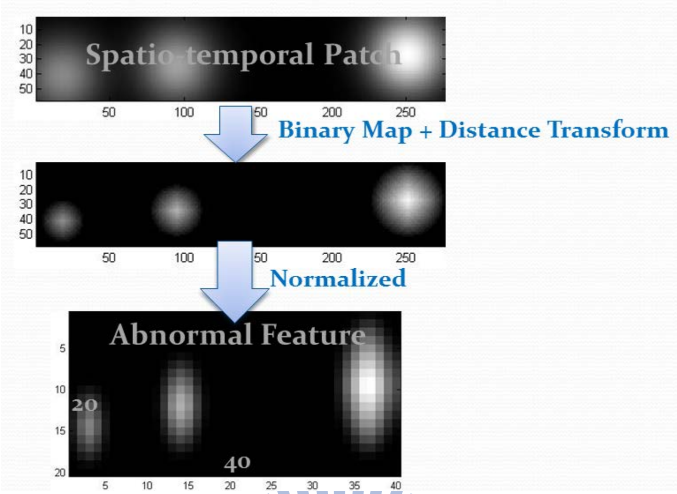

Figure 3-16 *Illustration of abnormal feature normalization*

In our work, we defined three usual behaviors. They are running, walking and stopping. If the predicted value does not fall into one of the labeled value, we view this behavior as an abnormal behavior. We have to emphasis that although the patch of running behavior is labeled and trained beforehand. We still take running behavior as an abnormal behavior since we are interested in the reason of running.

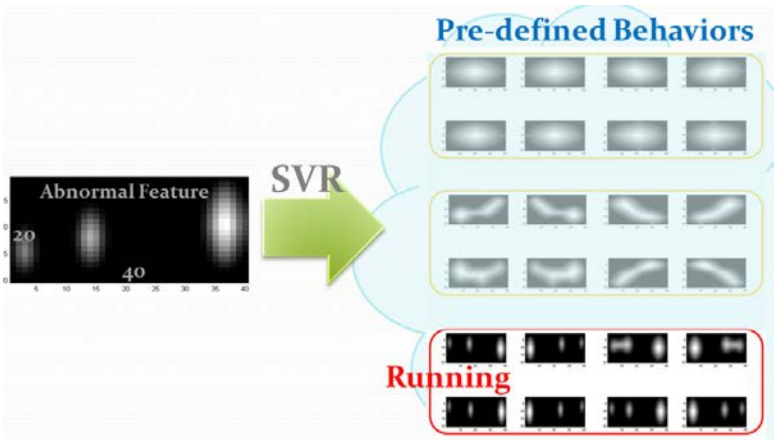

Figure 3-17 *Illustration of abnormal feature matching*

Chapter 4.

EXPERIMENTAL RESULTS

In this chapter, we will present our experimental results. To demonstrate our detected results, we have to briefly describe our surveillance environment. Our surveillance videos are taped outdoors from a bird-eye view. The frame rate of our surveillance videos is 15 fps and the resolution of our surveillance videos is 600 × 800. Since our proposed method is based on the prediction step, we do not process every consecutive frame. We take 3 frames in 1 second to observe pedestrians instead. That is, we observe pedestrian once in every 5 frames. We have to emphasis that the trade-off exists between pedestrian tracking and abnormal behavior detection. Pedestrian tracking always want to work at low frame rate. However, to detect abnormal behaviors accurately, we need more observations to improve the correctness and we take 3 frames in 1 second to balance these two in our surveillance system.

Our processing information is detailed above, then, we are able to present our experimental results. The rest of this chapter is organized as following: First of all, we will present the accuracy of constructing our module of pedestrian behavior modeling. Then, we will demonstrate some abnormal behaviors which can be detected both by using our proposed method and by constructing a scene model. Then, we will show some behaviors which will be viewed as abnormal behaviors but those behaviors will not be taken as abnormal behaviors in our proposed method. Finally, we will present some false alarm of abnormal behavior in our method.

As we mentioned before, we have five parameters to be determined in training step. The value of these five parameters is shown in table 4-1.

Table 4-1 Model parameters

σ_s	σ_g	w_{ori}	w_g	w_{cv}	Least Square Error
14	17.5	0.444	4	6.667	1250

We take another normal surveillance video in 1.5 minute to test our constructed model compare with linear prediction. The result is listed in table 4-2. We can find out that our proposed model is about 3.5 better than linear prediction.

Table 4-2 Comparison with linear prediction

	Least Square Error
Our proposed method	18486
Linear prediction	65883

The other result is the accuracy of group prediction. We take 2575 pair-wise features for training data and we take another 200 pair-wise features for testing. We got about 96.6% accuracy of group prediction with our proposed pair-wise feature.

As we mentioned earlier in our unusual behavior detection, we defined some pre-defined normal behaviors in our data set. The pre-defined behaviors in our data set are stopping, walking and running. As we recognize the detected behavior belongs to one of these pre-defined behaviors. This detected behavior will not be taken as an abnormal behavior and it will be viewed as warning instead. We draw yellow box as our warning behaviors. However, if the detected behavior does not belong to the pre-defined behavior, it will be viewed as an abnormal behavior and we use red box to present our detected abnormal behaviors.

There is one exception that if the detected behavior is recognized as accelerate suddenly, it will also be taken as an abnormal behavior and we use red box to present this behavior too.

At the following discussion, we will demonstrate our results of abnormal behavior detection. One of the easiest abnormal behaviors is the pedestrian accelerate suddenly as shown in Figure 4-1.

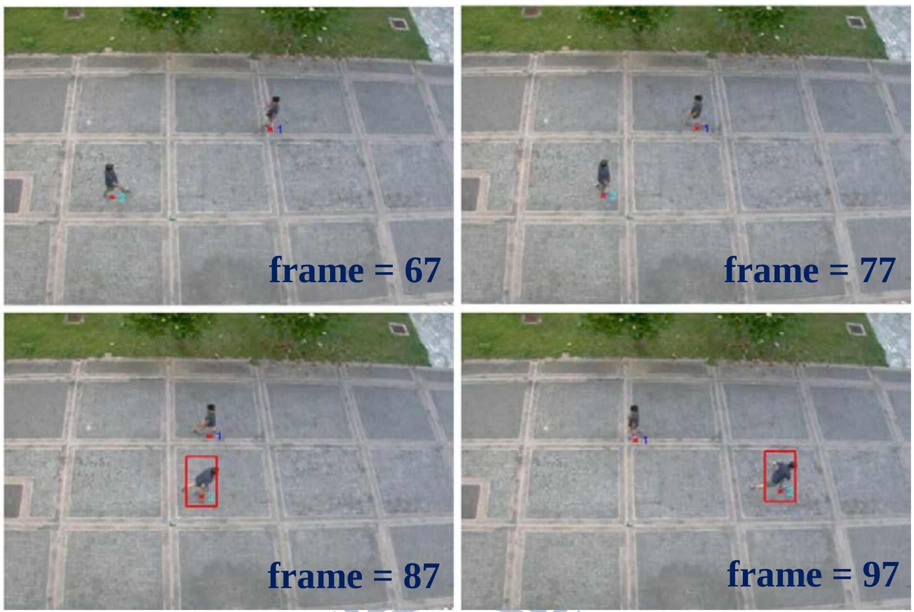

Figure 4-1 *Sudden acceleration*

Another easy abnormal behavior is the pedestrian walks into restricted region. In our case, the pedestrian walks onto grass region which is defined as a restricted in our scene as shown in Figure 4-2.

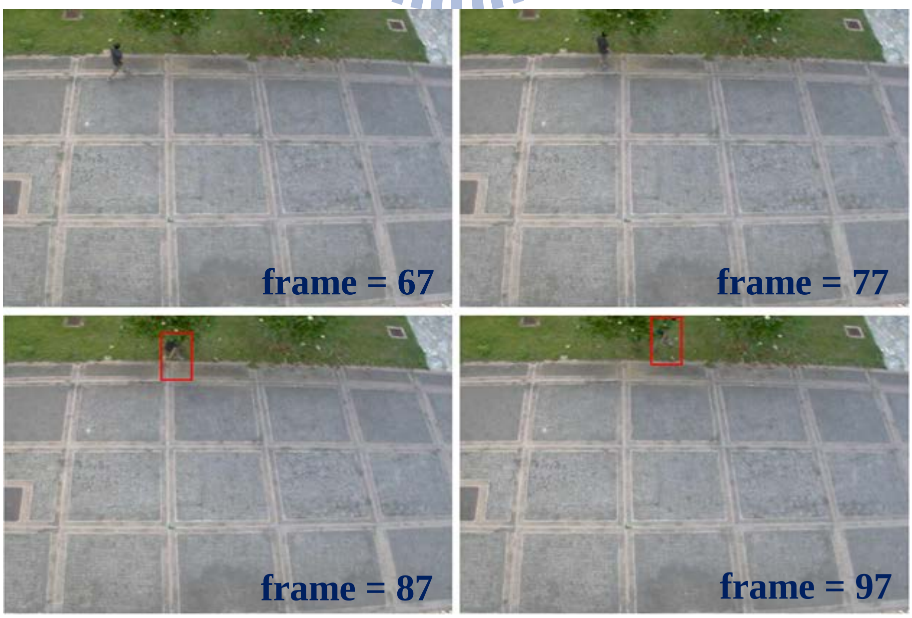

Figure 4-2 *Walk onto grass region*

The next abnormal behavior is robbery as shown in Figure 4-3. The concept of robbery requires intelligent knowledge and this action cannot be recognized in our surveillance system. However, since robbery mainly contains simple actions such as suddenly running and these simple behaviors can be detected in our proposed method as shown in Figure 4-3.

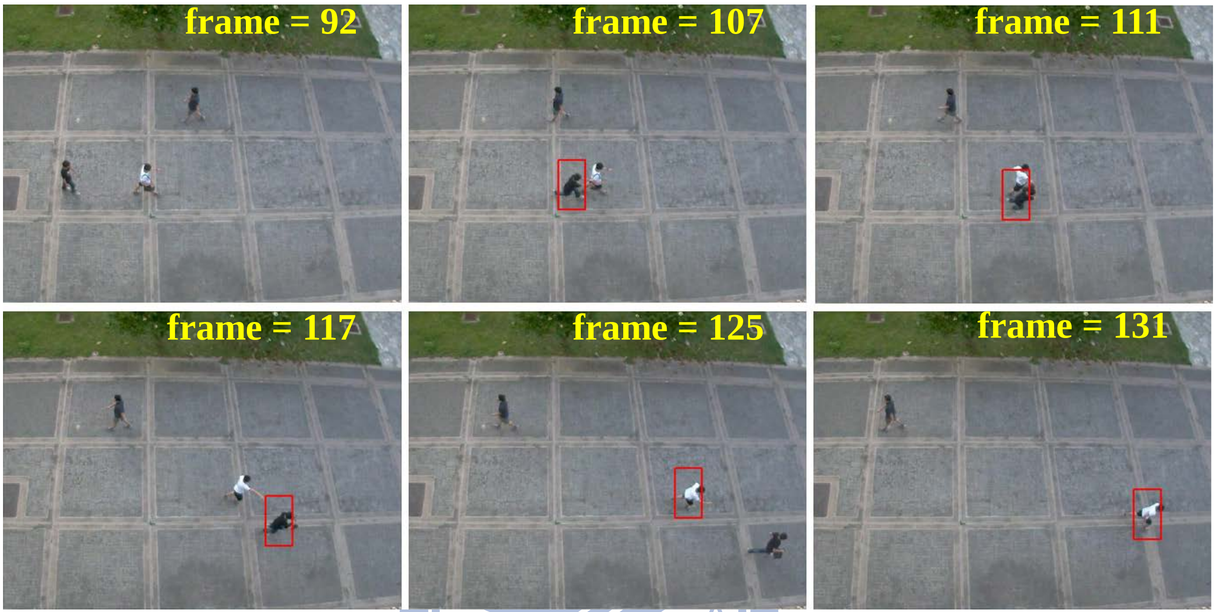

Figure 4-3 *Robbery*

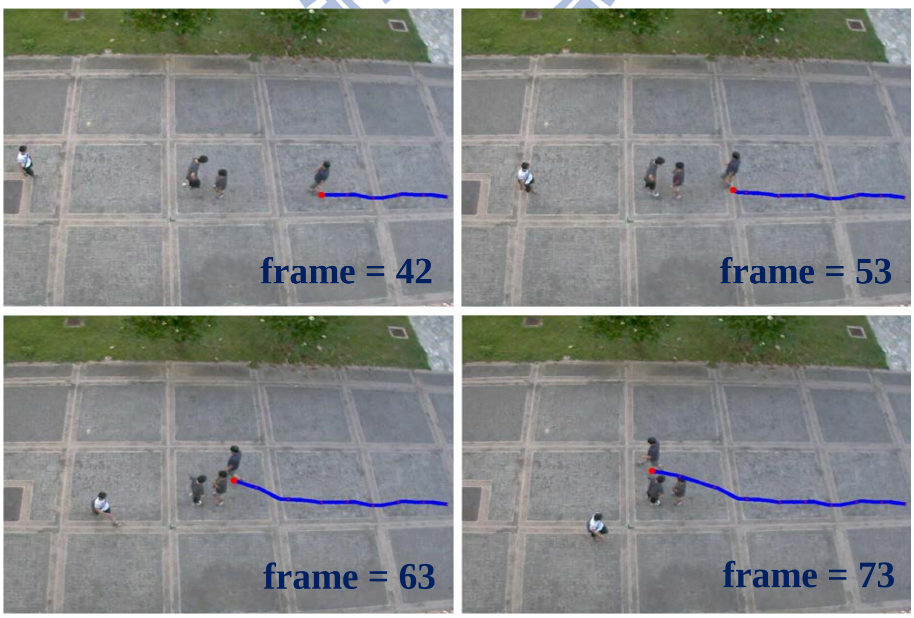

Figure 4-4 *Avoidance phenomenon without warning*

These three abnormal behaviors above can be easily detected in most of surveillance systems. The following examples show that our proposed method is insensitive to the avoidance phenomenon. The avoidance behavior might be viewed as abnormality in most of surveillance systems since the trajectory of avoidance phenomenon is different from normal trajectory. These examples of avoidance behavior are shown in Figure 4-4 and in Figure 4-5.

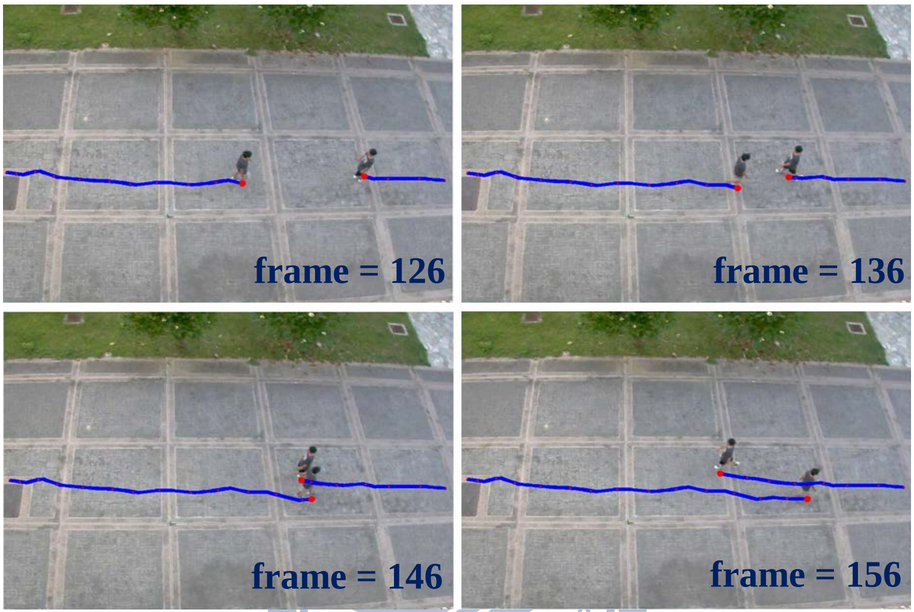

Figure 4-5 *Avoidance phenomenon without warning*

Another new example behavior is shown below in Figure 4-6.

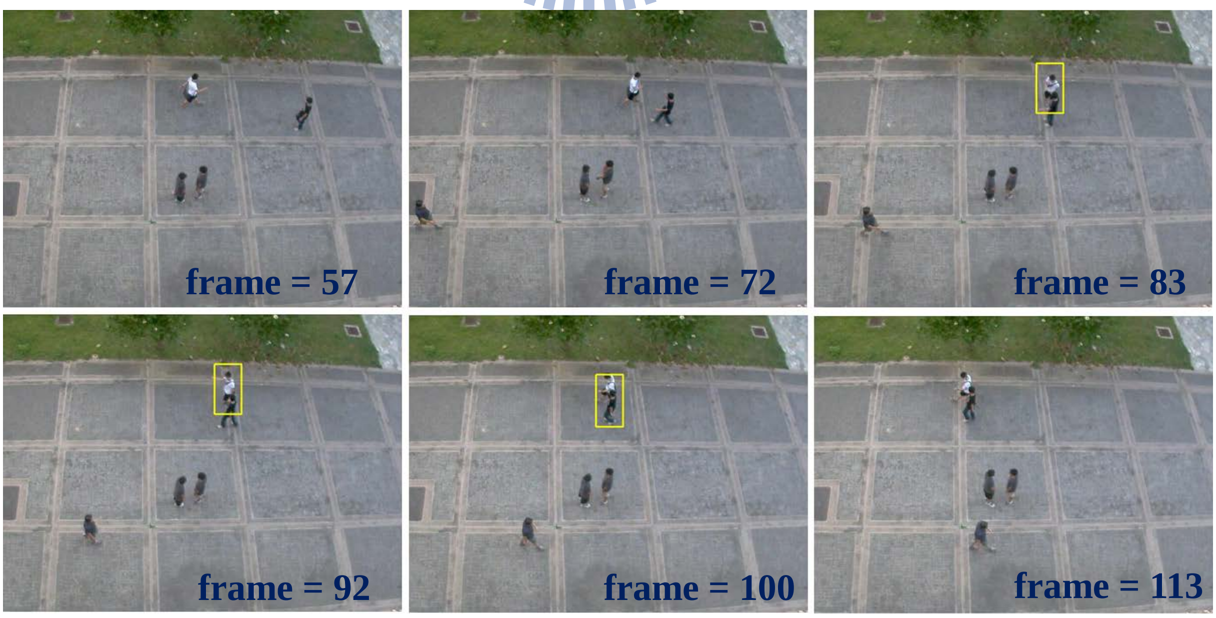

Figure 4-6 *Greet and walk together with his friend*

It presents that a pedestrian greet with his friend at the center of scene. Then, this pedestrian accompanies his friend toward the direction which is opposite to his original direction. In this case, the action of greeting can be captured at frame 83 and recognized as a stopping action. As he turns around toward opposite direction, this movement can be captured at frame 92 and recognized as a walking action. The status of social group for these two pedestrian has been changed since frame 92, the changing status of social group for these two pedestrian is also be captured at frame 100.

The next example is two pedestrians start from different sides and meet at the center of the scene. They chat for a while and move separately toward their original destinations as shown in Figure 4-7.

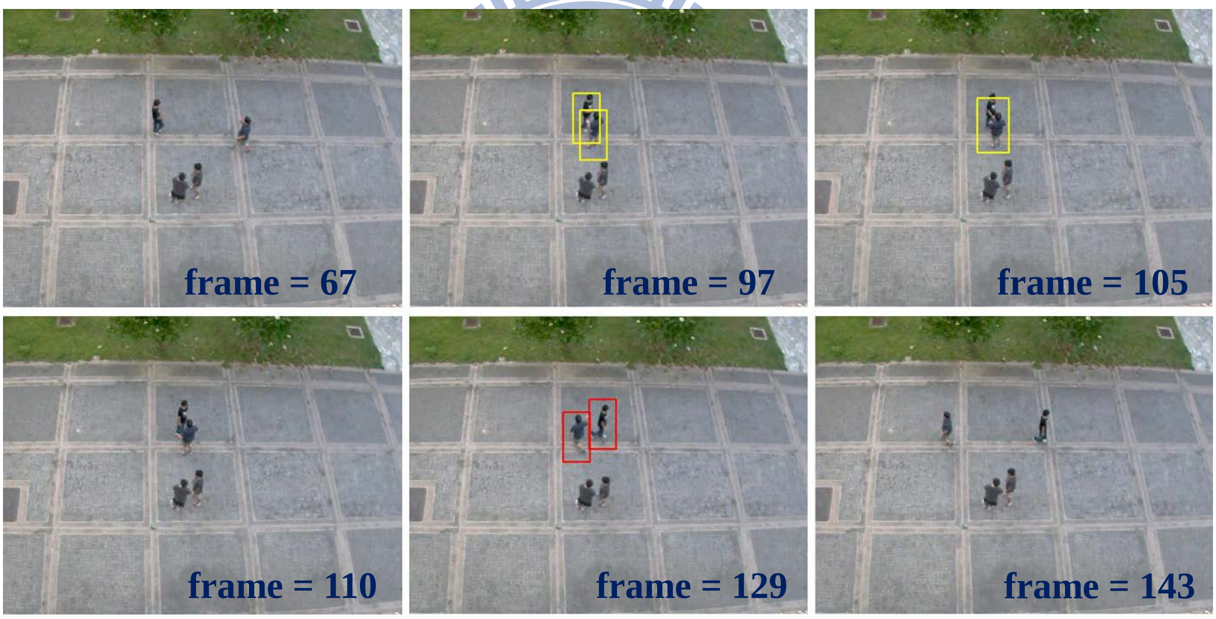

Figure 4-7 *Greet and chat with friend, then move separately*

As they meet at the center of scene at frame 97, the behavior of stopping is detected for each pedestrian. As they stay together longer, they will be viewed as a group as shown at frame 105. Then the group status will maintain until they separate at frame 129 and this behavior of separation will be recognized as an abnormal behavior.

The next example shows that friends toward different destinations as shown in

Figure 4-8. As they separate at frame 72, we are able to detect this behavior since the group status is changed.

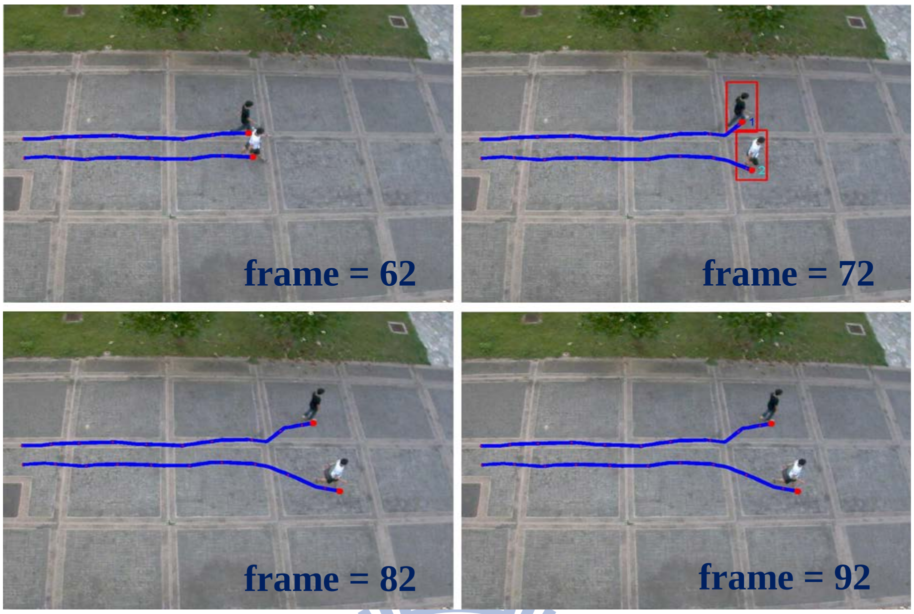

Figure 4-8 *Friends toward different destinations*

The next example shows that friends group together as shown in Figure 4-9.

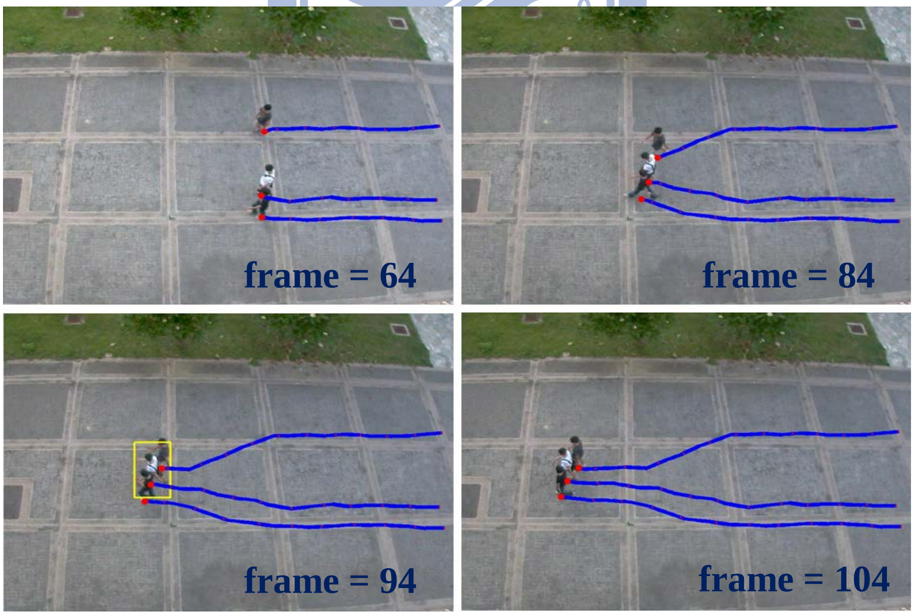

Figure 4-9 *Friends group together*

As they group together at frame 94, we are able to detect this behavior since the group status is changed.

However, these are still some unwanted warnings in our surveillance system as shown in Figure 4-10.

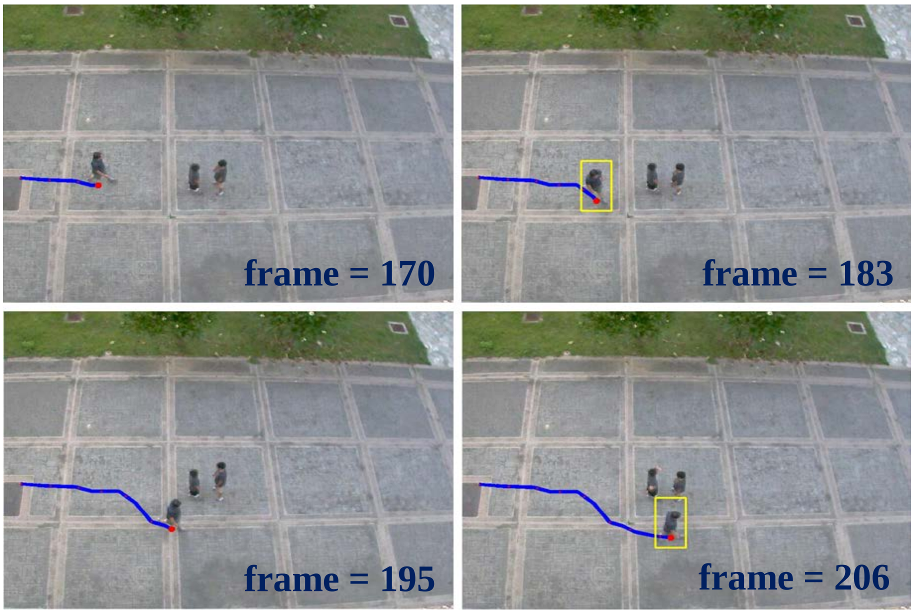

Figure 4-10 *Avoidance phenomenon with warning*

It is because the avoidance phenomenon of pedestrian is much more serious than we expected and we are not able to handle this movement even it just an avoidance action.

Chapter 5.

CONCLUSIONS

In this thesis, we introduce a module of pedestrian behavior modeling which tries to model pedestrian behavior into our tracking scheme. This module of pedestrian behavior modeling takes social interactions between objects and social interactions between scene and objects into account to model possible behaviors of pedestrian. By introducing this module, we track pedestrian from the pedestrian aspect which are quite different from most tracking methods. Most of tracking methods do not take the social factors which effects behavior of pedestrian heavily into consideration.

As for the abnormal behavior analysis, we detect abnormal behaviors based on the introduced module rather than constructing a normal scene model. Since the module of pedestrian behavior modeling is from the aspect of pedestrian, we are able to detect much more abnormalities by using the introduced module of pedestrian behavior modeling than by using a scene model which is constructed in most approaches for abnormal behavior detection.

REFERENCES

[1] A. Yilmaz, O. Javed and M. Shah, "Object Tracking: A Survey", *ACM Computing Surveys*, vol. 38, no. 4, article 13, December 2006.

[2] Y. Bar-Shalom and T. Fortmann, "Tracking and Data Association", *Academic Press*, 1988.

[3] I. Cox, "A Review of Statistical Data Association Techniques for Motion Correspondence", *International Journal on Computer Vision*, vol. 10, no. 1, pp. 53-66, 1993.

[4] A. Basharat, A. Gritai and M. Shah, "Learning Object Motion Patterns for Anomaly Detection and Improved Object Detection", *IEEE conference on Computer Vision and Pattern Recognition*, pp. 1-8, January 2008.

[5] V. Mahadevan, W. Li, V. Bhalodia and N. Vasconcelos, "Anomaly Detection in Crowded Scenes", *IEEE conference on Computer Vision and Pattern Recognition*, pp. 1975-1981, 2010.

[6] A. Treuille, S. Cooper, Z. Popovi´c, "Continuum Crowds", *ACM SIGGRAPH*, vol. 25, no.3, pp. 1160-1168, 2006.

[7] A. Lerner, Y. Chrysanthou and D. Lischinsk, "Crowds by Example", *EUROGRAPHICS*, vol. 26, no. 3, pp. 655-664, September 2007.

[8] D. Helbing and P. Molnar, "Social Force Model for Pedestrian Dynamics", *Physical Review E*, vol. 51, no. 5, May 1995.

[9] D. Helbing and T. Vicsek, "Simulating Dynamical Features of Escape Panic", *Nature*, vol. 407, no. 6804, pp. 487-490, 2000.

[10] G. Antonini, S. Martinez, M. Bierlaire and J. Thiran, "Behavioral Priors for Detection and Tracking Pedestrian in Video Sequences", *International Journal of Computer Vision*, vol. 69, no. 2, pp. 159-180, August 2006.

[11] S. Pellegrinin, A. Ess, K. Schindler, L. van Gool, "You'll Never Walk Alone: Modeling Social Behavior for Multi-target Tracking", *IEEE International Conference on Computer Vision*, pp. 261-268, 2009

[12] P. Scovanner and M. Tappen, "Learning Pedestrian Dynamics from The Real World", *IEEE International Conference on Computer Vision*, pp. 381-388, 2009

[13] A. Hoogs, S. Bush and G. Brooksby, "Detecting Semantic Group Activities Using Relational Clustering", *IEEE Workshop on Motion and Video Computing*, pp. 1-8, January 2008.

[14] W. Ge, R. Collins and B. Ruback, "Automatically Detecting the Small Group Structure of a crowd", *IEEE Workshop on Applications of Computer Vision*, pp. 1-8, December 2009.

[15] V. Pavlovic, J. M. Rehg and J. MacCormick, “Learning Switching Linear Models of Human Motion”, *Neural Information Processing Systems*, pp. 981-987, 2000