DETECTION AND CLASSIFICATION OF MOVING OBJECTS FOR AN AUTOMATED SURVEILLANCE SYSTEM

[illegible]

MASTER OF SCIENCE
UNIVERSITI PUTRA MALAYSIA

2006

DETECTION AND CLASSIFICATION OF MOVING OBJECTS FOR AN AUTOMATED SURVEILLANCE SYSTEM

By

MOHD RAZALI BIN MD TOMARI

Thesis Submitted to the School of Graduate Studies, Universiti Putra Malaysia,
in Fulfilment of the Requirement for the Degree of Master of Science

September 2006

Dedicated to my loving family, for their endless support

Abstract of thesis presented to the Senate of Universiti Putra Malaysia in fulfilment of the requirement for the degree of Master of Science

DETECTION AND CLASSIFICATION OF MOVING OBJECTS FOR AN AUTOMATED SURVEILLANCE SYSTEM

By

MOHD RAZALI BIN MD TOMARI

September 2006

Chairman: **Associate Professor Adznan Jantan, PhD**

Faculty: **Engineering**

Automated surveillance system has been the subject of much research recently. A completely automated system means a computer will perform the entire task from low level detection to higher level motion analysis. Since conventional system practically using human power to monitor and did not applicable for a long hour monitoring, thus automated system had been created to replace the conventional system. This thesis focuses on a method to detect and classify a moving object that pass through the surveillance area boundary. Moving object is detected by using combination of two frame differencing and adaptive image averaging with selectivity. Technically, this method estimate the motion area before updates the background by taking a weighted average of non-motion area of the current background altogether with non-motion area of the current frame of the video sequence. This step had created a focus of attention for higher level processing and it helps to decrease computation time considerably. The output of a motion–based detector is essentially a collection of foreground that might correspond to the moving objects. But usually the output image produced from this process contaminated with noise and shadow. As a solution, morphological operation has been employed as an approach to remove noise from the foreground object. Mutual shadow that exists with the object had been abolished by combining chromatic colour values with lightness variable. Then, standardized moment invariant is employed to extract the features for each moving blobs. To recognize these blobs, the calculated moment values are fed to a support vector machine module that is equipped with trained extracted moment values for human and vehicle silhouettes. The system operates on colour video imagery from a stationary camera. It can handle object detection in outdoor environments and under changing illumination conditions. The applied post processing module capable to remove noise and shadow from the detected objects with less than 1% of error. Finally, classification algorithm that makes use of the extracted moment values from the detected objects successfully categorize objects into pre-defined classes of human and vehicle with 89.08% of accuracy. All the methods have been tested on video data and the experimental results have demonstrated a fast and robust system

Abstrak tesis yang dikemukakan kepada Senat Universiti Putra Malaysia sebagai memenuhi keperluan untuk ijazah Master Sains

PENGESANAN DAN KLASIFIKASI OBJEK-OBJEK BERGERAK UNTUK SISTEM PENGAWASAN AUTOMATIK

Oleh

MOHD RAZALI BIN MD TOMARI

September 2006

Pengerusi: **Profesor Madya Adznan Bin Jantan, PhD**

Fakulti: **Kejuruteraan**

Sistem pengawasan automatik telah menjadi antara bidang penyelidikan yang utama ketika ini. Sistem pengawasan automatik menyeluruh bermakna komputer melakukan semua kerja daripada peringkat terendah pengesanan hingga ke peringkat tinggi analisa pergerakan. Oleh kerana sistem sedia ada menggunakan manusia , ia tidak sesuai dan berkualiti untuk pegawasan dalam tempoh waktu yang lama, maka sistem automatik ini merupakan alternatif terbaik menggantikan sistem konvensional tersebut. Tesis ini memfokuskan kaedah untuk mengesan dan mengklasifikasi objek bergerak yang merentasi kawasan pengawasan. Objek bergerak dikesan menggunakan kombinasi teknik pembezaan dua kerangka dan teknik purata imej suai dengan pemilihan, dimana secara teknikalnya, kaedah ini menganggar kawasan pergerakan sebelum mengemaskini latarbelakang dengan mengambil kira purata piksel pemberat diluar kawasan pergerakan daripada latarbelakang dan kerangka terkini daripada susunan video. Langkah ini memfokuskan kepada kawasan yang lebih khusus dan kecil untuk proses yang lebih tinggi ,dengan itu secara tidak langsung mengurangkan masa untuk pengiraan. Hasil daripada pengesan pergerakan ini ialah koleksi penting latar depan yang merupakan objek bergerak. Namun biasanya hasil imej daripada proses ini dicemari dengan hingar dan bayang-bayang. Sebagai langkah penyelesaiannya, operasi morfologi dipilih sebagai cara untuk membersihkan hingar daripada objek latar depan. Bayang-bayang yang terdapat pada objek pula dihapuskan dengan kombinasi nilai warna kromatik dan pembolehubah cahaya. Selepas itu piawaian momen tak varian digunakan untuk mengekstrak ciri daripada objek bergerak. Untuk mengecam objek ini, nilai momen yang telah dikira dihantar ke modul mesin penyokong vektor yang sebelum itu dilengkapkan dengan pemahaman tentang ekstrak nilai momen daripada bebayang bentuk manusia dan kenderaan. Sistem ini beroperasi menggunakan video warna daripada kamera yang dalam keadaan pegun. Ia boleh mengesan objek di persekitaran luar dan dalam keadaan perubahan keamataan cahaya. Modul pemprosesan pasca mampu menghapuskan hingar dan bayang-bayang daripada objek yang dikesan dengan ralat kurang daripada 1%. Akhir sekali, algoritma pengelasan menggunakan nilai momen yang telah diekstrak daripada objek yang dikesan berjaya mengkategorikan objek samada manusia atau kenderaan dengan ketepatan 89.08%. Semua kaedah ini telah diuji pada data video dan keputusan eksperimen membuktikan bahawa sistem ini pantas dan tegap.

ACKNOWLEDGEMENTS

I would like to express my gratitude to Assist. Prof. Dr. Adznan Jantan for his supervision, encouragement, suggestions and trust throughout the development of this thesis.

I owe special thanks to Dr. Khairi bin Yusuf for his guidance, support and invaluable discussions that encouraged me in my research.

I am also thankful to my research group members. With their support my research becomes more enjoyable for me. Not to forget, my biggest gratitude goes my family for their endless love, support and trust in me. Without them I would never come up to this stage.

Finally, although my name is officially printed in this thesis, the contribution is come from all of you. So I specially dedicate this thesis for you.

I certify that an Examination Committee has met on 14 September 2006 to conduct the final examination of Mohd Razali Bin Md Tomari on his Master of Science thesis entitled "Detection and Classification of Moving Objects for an Automated Surveillance System" in accordance with Universiti Pertanian Malaysia (Higher Degree) Act 1980 and Universiti Pertanian Malaysia (Higher Degree) Regulations 1981. The Committee recommends that the candidate be awarded the relevant degree. Members of the Examination Committee are as follows:

Ir. Wan Ishak Wan Ismail, PhD
Professor
Faculty of Engineering
Universiti Putra Malaysia
(Chairman)

Mohd Adzir Mahadi, PhD
Associate Professor
Faculty of Engineering
Universiti Putra Malaysia
(Internal Examiner)

Abdul Rahman Ramli, PhD
Associate Professor
Faculty of Engineering
Universiti Putra Malaysia
(Internal Examiner)

Che Mat Hadzer Mahmud, PhD
Associate Professor
School of Electrical System
Kolej Universiti Kejuruteraan Utara Malaysia
(External Examiner)

HASANAH MOHD. GHAZALI, PhD
Professor/ Deputy Dean
School of Graduate Studies
Universiti Putra Malaysia

Date: 21 DECEMBER 2006

This thesis submitted to the Senate of Universiti Putra Malaysia and has been accepted as a fulfilment of the requirement for the degree of Master of Science. The members of the Supervisory Committee are as follows:

Adznan Jantan, PhD
Associate Professor
Faculty of Engineering
Universiti Putra Malaysia
(Chairman)

Khairi Yusuf, PhD
Lecturer
Faculty of Engineering
Universiti Putra Malaysia
(Member)

AINI IDERIS, PhD
Professor / Dean
School of Graduate Studies
Universiti Putra Malaysia

Date: 16 JANUARY 2007

DECLARATION

I hereby declare that the thesis is based on my original work except for quotations and citations which have been duly acknowledged. I also declare that it has not been previously or concurrently submitted for any other degree at UPM or other institutions

MOHD RAZALI BIN MD TOMARI

Date:20 DECEMBER 2006

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

AOF	Average Optical Flow
AVI	Audio Video Interleave
CMU	Carnegie Mellon University
CCL	Connected Component Labelling
COM	Centre of Mass
dll	Dynamic Link Library
DDD	Dilation-Dilation-Dilation
E	Erosion
EEE	Erosion-Erosion-Erosion
FD	Fourier Descriptor
FC	Fourier Coefficients
FP	False Positive
FN	False Negative
GRBF	Gaussian Radial Basis Function
GUI	Graphic User Interface
HSV	Hue, Saturation, Value
IIR	Infinite Impulse Response
LBP	Local Binary Pattern
LDA	Linear Discriminant Analysis
MIT	Massachusetts Institute of Technology
MoG	Mixture of Gaussians
Matlab	Matrix Laboratory

MLP	Multilayer Perceptron
MBBR	Motion Bounding Box Region
PCA	Principal Component Analysis
QP	Quadratic Programmimg
RGB	Red, Green, Blue
ROG	Radii of Gyration
ROI	Region of Interest
SMM	Shading Model Method
SBA	Selective Background Updating with Averaging
SV	Support Vector
SSE	Streaming Single Instruction Multiple Data Extension
SVM	Support Vector Machine
TP	True Positive
TN	True Negative
VSAM	Video Surveillance and Monitoring
VC	Vapnik Chervonenkis
2D	Two Dimensions
1D	One Dimension

CHAPTER I

INTRODUCTION

In recent years, with huge evolution and advancement in computer world, intelligent vision has become an active area of research , with the goal of developing visual sensing as well as processing algorithms and hardware that can distinguish and understand the world around them. Among those, visual surveillance system receives a great deal of interest. Video surveillance has been applied widely to ensure better precautions in security-sensitive areas, like factory, airports, schools or government offices.

Traditionally, the most important task of monitoring precautions is primarily based on human visual observation, which is a hard work for watchmen. During a long hour of monitoring, human concentration will slightly decrease and simultaneously affect the efficiency of the monitoring system. In addition, area enclosed under surveillance may be too large to be monitored by a few operators whereas number of cameras might exceed their monitoring capability.

These problems urge the usage of automation in surveillance system where computer performs the task that human operator normally would. Vast amount of data acquired from video imagery will be analyzed by an intelligent and useful autonomous structure. Also, this intelligent system will have the capacity to observe the surrounding environment and extract useful information for subsequent reasoning, such as detecting and analyzing the activity (motion), or identifying objects that enter the scene. Even though this system cannot completely replace the human's presence, it will provide a great help for the watchmen to monitor large surveillance area with minimum human power supervision.

The formation of intelligence surveillance systems requires fast, reliable and robust algorithms for moving object detection, classification, and activity analysis. Moving object detection is the first step towards activity analysis. Commonly used techniques for this purpose are background subtraction, temporal differencing and optical flow [1]. This step not only creates a focus of attention for higher level processing but also decreases computation time considerably. The output of a motion–based detector is an essential collection of foreground that might correspond to the moving objects. However, classification of these regions into different categories of objects is still a huge challenge.

Object classification step categorizes detected objects into predefined classes such as human, vehicle, animal, etc. It is necessary to distinguish objects from each other in order to analyze their reliable actions. Currently, there are two major approaches towards moving object classification, which are shape-based and motion-based methods [2]. Shape-based methods make use of the objects' two dimensional spatial information whereas motion-based methods use temporal tracked features of objects for the classification solution.

Both, the outputs detection and classification algorithms can be used for providing the human operator with high level data in order to yield accurate decisions within a short time besides offering an effective offline indexing practice and a proficient routine to search for stored video data. Advancement in the development of these algorithms would lead to breakthroughs on applications that use visual surveillance. Table 1.1 showed some scenarios that these algorithms might handle [5] [7] [8] [9] [10] [13] [15] [19] [22] [25] [48]:

Table 1.1: Automated visual surveillance system application.

Application area	Example of the application
Public and commercial security	i. Monitoring banks, department stores and parking lots. ii. Patrolling highways and railways for accident detection. iii. Access control
Smart video data mining	i. Measuring traffic flow and pedestrian congestion. ii. Counting vehicle that entering and leaving the scene.
Law enforcement	i. Measuring the speed of vehicles ii. Detecting red light crossings and unnecessary lane occupation

1.1 Objectives of Research

Automated surveillance system carries a large number of benefits especially for safety precaution. The objectives of this research are:

i. To develop a motion detector module that can robustly detect and segment motions accurately from captured video sequences, in RGB colour mode and capable to cope with the changes in the scene.
ii. To propose a method for eliminating noise and shadow, from the segmented blobs, and extract important features for classification determination.
iii. To develop algorithm of an object classification system that employs the filtered blobs based on supervised learning with a small number of labelled examples, to distinguish between human and vehicle.

The software is developed using C++ and Visual Basic.

1.2 Scope of Thesis

This thesis deals with the problems of defining and developing the building blocks of moving object detection and classification system. The scope of this thesis is on method to detect and distinguish semantically-different classes of objects which have gross differences. The system can perform the classification task for multiple objects as long as the object is not occluded. Besides, it is limited to classify between human and vehicle class, for video inputs from static camera where the view frustum that may change arbitrarily are not supported. The corresponding performances of the proposed system blocks are validated by examine the extent of similarity between the outputs from the classified image with the ground truth.

1.3 Thesis Outline

This thesis is being divided into five consecutive chapters where each chapter reviews different issues regarding to the project objectives. Chapter 1 covers the introductory section of the project while Chapter 2 describes the literature review and theoretical background that related to automated surveillance system. The following Chapter 3 provides the explanation on project methodology used throughout the operation of the project analysis, result, and discussion are explained individually in Chapter 4 and the last chapter, which is Chapter 5, considers the conclusion and future recommendations in extending the project into a better prospect.

CHAPTER 2

LITERATURE REVIEW

The framework described in this study includes four building blocks for an automated surveillance system, which can be listed as moving object detection, post processing, feature extraction and object classification. This chapter describes the theory, related work and studies in the literature on each of these building blocks and the complete system of automated surveillance system.

2.1 Moving Object Detection

Detecting changes in image sequences of the same scene, captured at different times, is of significant interest due to a large number of applications in several disciplines. Video surveillance is among the important applications, which requires reliable detection of changes in the scene. Detecting a region that correspond to the moving objects in video is the first basic step of almost every vision system since, it provides a focus of attention and simplifies the processing on subsequent analysis steps. Due to dynamic changes in natural scenes such as sudden illumination and weather changes, repetitive motions that cause clutter (tree leaves moving in blowing wind), motion detection is a difficult problem to process reliably. There are mainly three methods to detect moving objects which are optical flow, frame differencing and background subtraction [1].

2.1.1 Optical Flow

Optical flow is a velocity field in the image that transforms one image into the next image in a sequence [14]. It can be used to detect and track targets without only a prior knowledge about the background, this method can also be performed when the camera is moving. Optical flow is calculated based on the changes in the image intensity originate from local translation of the pixels. Many of the optical flow algorithms are proposed in the recent years, among of them are Horn and Schunck [14], Lucas and Kanade [15], Anandan [16], Nagel [17], Uras [18], Camus [19] and Proesmans [20]. Despite their differences, many of these techniques can be viewed conceptually in terms of three stages of processing [21]:

1. Pre filtering or smoothing with low pass or band pass filters in order to extract signal structure of interest and to enhance the signal-to-noise ratio.
2. The extraction of basic measurements, such as spatiotemporal derivatives (to measure normal components of velocity) or local correlation surfaces.
3. The integration of these measurements to produce a two dimension flow field, which often involves assumptions about the smoothness of the underlying flow field.

From the evaluation in [22], it concludes that Lucas and Kanade [15] method produced an accurate depth maps, simple assumption, and good noise tolerance. However, most of optical flow method are computationally complex and cannot be used in real-time without specialized hardware [3].

2.1.2 Frame Differencing

Frame differencing method attempts to detect moving regions by making use of the pixel-by-pixel differences of two consecutive frames in a video sequence. The implementation of this method in [23], [4] found that it can detect a moving object in real time and highly adaptive to the dynamic scene changes. However, it generally fails in detecting whole relevant pixels of some types of moving objects and fails to detect stopped objects in the scene. The common equation for detecting foreground using two frame differencing method is shown below [23]:

$$|I_t(x,y) - I_{t-1}(x,y)| > \tau \qquad (2.1)$$

where

τ is a threshold values.

$I_t(x, y)$ is pixel intensity at respected x and y coordinate

Additional methods need to be adopted in order to detect stopped objects for the success of higher level processing. In order to overcome shortcomings of two frame differencing in some cases, three frame differencing [5] were used. This improvement can obtain a better result than two frame differences, but the shortcomings of frame difference are keeping up because it cannot detect the stopped object and thus forces to turn up to the other detection method .

2.1.3 Background Subtraction

Background subtraction is a commonly used technique for segmenting out objects of interest in scene applications such as surveillance. It attempts to detect moving regions by subtracting the current image pixel-by-pixel from a reference background image that is created by averaging images over time in an initialization period [1]. The pixels where the difference is above a threshold are classified as foreground. The reference background is updated with new images over time to adapt to dynamic scene changes.

The block diagram of background subtraction is shown in Figure 2.1 consists of three main components which are background modelling, foreground detection, and post processing [6]. Background modelling uses the new video frame to calculate and update a background model. This background model provides a statistical description of the entire background scene. Foreground detection then identifies pixels in the video frame that cannot be adequately explained by the background model, and outputs them as a binary candidate foreground mask. Finally, post processing examines the candidate mask, eliminates those pixels that do not correspond to actual moving objects, and outputs the final foreground mask.

Most background subtraction algorithms detect changes in the current frame by comparing it with a reference frame, where the reference frame may be the previous frame or the estimated background. The techniques for performing the comparison can be divided into three categories [7]:

1. Feature Matching: These techniques attempt to match features, such as edges and corners, and detect differences in the location of the features.
2. Difference and Threshold: This is probably the most widely used type of technique. The difference between the current and reference frames is determined, and for each pixel if the value exceeds a threshold the pixel is flagged as having changed.
3. Statistical: In these techniques statistical properties over regions of the current and reference images are compared, and if the statistical property differ sufficiently then these regions are flagged as having changed.

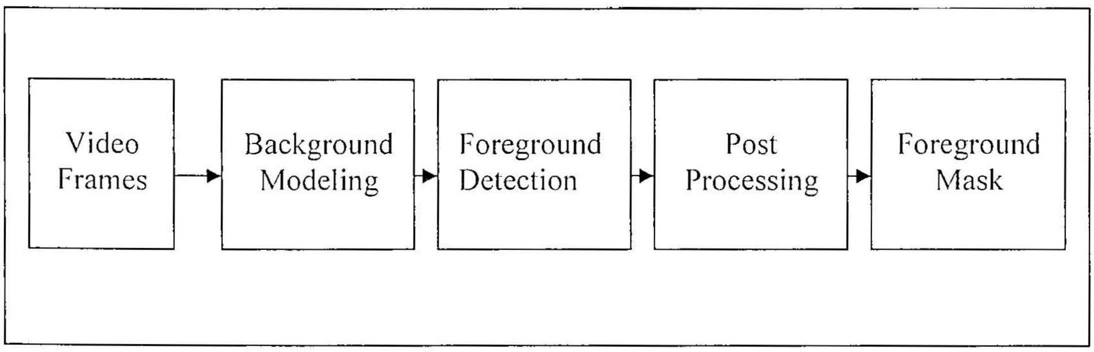

Figure 2.1: Block diagram of background subtraction algorithm

Background modelling is an important part of any background subtraction algorithm. Acquiring the background model is a complicated task. The most straightforward approach would be to simply set up the camera, empty the scene of any moving objects and take a snapshot. Although this approach is simple, it is always impractical in real scenes because background can change over time, it is difficult to empty a scene, lighting can change, and the camera position can drift. A good background model should adapt scene changes and need to be updated continuously. The updated is normally performed in a first in first out manner, means that the oldest sample is discarded, and a new sample is added to the model. There are two alternative mechanisms to update the background which listed below [24]:

1. Selective update: Add the new sample image to the model only if it is classified as background sample. This method enhances detection of the targets, since target pixels are not added to the model. The simplest way to do this is to use the detection result as an update decision. The problem of this approach is that any incorrect detection decision will result an incorrect detection later which cause a deadlock situation [25]. For example, if one object is added to the background scene and stay for a long time, it will be continually detected as a foreground object.
2. Blind update: Add the entire sample image to the model. This approach does not suffer from the deadlock situation since it does not involve any update decision and allows intensity values that do not belong to the background to be added to the model. This lead to bad detection of the targets as they become part of the model. This effect is reduced when the time window is increased, but more misclassified occurs because the adaptation to changes is slower.

Much research has been devoted to enhance the updated method of a background model to make it robust against environmental changes in the background, but sensitive enough to identify all moving objects of interest. The comprehensive comparison and method of background modelling techniques can be found in [1] [7] [8] [13]. The model can be classified into two broad categories which are non recursive and recursive techniques. Some of the methods will be described in the following subsections.

2.1.3.1 Non Recursive Techniques

A non recursive technique uses a sliding window approach for background estimation. It stores a buffer of the previous N video frames, and estimates the background image based on the temporal variation of each pixel within the buffer. Non recursive techniques are highly adaptive as they do not depend on the history beyond those frames stored in the buffer. On the other hand, the storage requirement can be significant if a large buffer is needed to cope with slow-moving traffic. Some of the commonly used non recursive techniques are described in the next paragraph.

Cucchiara, et al. [9] used median filtering to model the background from incoming input frame. The background is estimate based on the median at each pixel location of all the frames in the buffer. The assumption is that the pixel stays in the background for more than half of the frames in the buffer. They conclude that median function has proven effectiveness while at the same time, less computational cost than the Gaussian or other complex statistics. For updating the background, the selective background update was used. In order to avoid a deadlock, an average optical flow (AOF) is calculated for each object to decide whether the object should be added or eliminated from the background model. In order to detect the moving object, colour distance measure was implemented, in practical this method finds the maximum value between input and background image for each of the three RGB channel. If the maximum value is greater than the threshold value the pixel will be marked as move.

Toyama, et al. [27] described an algorithm called Wallflower that used a wiener filter to predict a pixel's current value from a linear combination of its N previous values in the buffer. Pixels whose prediction error is several times worse than the expected error are classified as moved pixel. The predictive coefficients are adaptively updated at each frame. The Wallflower algorithm also tries to correctly classify the interiors of homogeneously coloured moving object by determining the histogram of connected component of change pixels and adding pixels to the change mask based on distance and colour similarity. Wallflower attempts to solve many of the common problems with background maintenance such as handling stopping object, detect foreground object that have similarity with background and adapt to the changing in the environment.

Elgammal, et al. [24] use the entire of N history of image sample in buffer, to form a non-parametric estimate of the pixel probability density function using kernel estimator which was chosen to be Gaussian. For updating the background they used both selective and blind update. Selective update is used to update the short term model that adapt to quick changes in the scene, while the blind update is used to update a long term model that adapts to change slowly. The intersection between these two models that satisfy the threshold value is then used to update the background model. All current pixels then are declared as foreground if the probability value between current and background image is smaller than the predefined threshold. The advantage of estimating intensity directly from sample history values is the ability to handle multi-modal background distribution. Examples of multi-modal background include pixels from a swinging tree or near high-contrast edges where they flicker under small camera movement. As a matter of fact, this theoretically well established method yields many accurate results under challenging outdoor conditions such as snow and fog [24].

2.1.3.2 Recursive Techniques

Recursive techniques do not maintain a buffer for background estimation. Instead, they recursively update a single background model based on each input frame. As a result, input frames from distant past could have an effect on the current background model. Compared with non-recursive techniques, recursive techniques require less storage, but any error in the background model can affect the detection process for a much longer period of time. They are many research devoted to background modelling based on this technique. Some of the representative method techniques are described in the following paragraph.

Stauffer and Grimson [28] described an adaptive background mixture model for real-time tracking. In their work, every pixel is separately modelled by a mixture of Gaussians (MoG) which are updated online by incoming image data. The idea is to correctly classify dynamic background pixels, such as the swaying branches of a tree or the ripples of water on a lake. Every pixel value is compared against the existing set of models at that location to find match. The parameters for the matched model are updated based on a learning factor. If there is no match, the least-likely model is discarded and replaced by a new Gaussian with statistics initialized by the current pixel value. In order to detect the foreground and background of the new pixel, the Gaussian distributions are sorted according to a simple metric. The accumulated pixels define the background Gaussian distribution whereas scattered pixels are classified as foreground.

Block based approaches have been also used for modelling the background. Block matching has been used to detect change between two consecutive frames. In [29], both reference and current frame are divided evenly into a number of non overlapping blocks. Then the average grey level of pixel in each block is computed. Moving object is detected then by threshold the difference curve of the block average grey level of the two images. In [10], the method divides each new video frame into equally sized block by using partially overlapping grid structure. Each block then was modelled by a group of weighted local binary pattern (LBP) histogram. Subsequently, at each new frame, each block histogram is compared against background model histogram using histogram intersection distance measure. Any intersection values that exceed the given threshold values are considered to be foreground. The major drawback with block-based approaches is that the detection unit is a whole image block and therefore they are only suitable for coarse detection.

Edge features have also been used to model the background. The use of edge features to model the background is motivated by the desire to have a representation of the scene background that is invariant to illumination changes. In [30], foreground edges are detected by comparing the edges in each new frame with an edge map of the background which is called the background "*primal sketch*." The major drawback of using edge features to model the background is that it would only be possible to detect edges of foreground objects instead of the dense connected regions that result from pixel-intensity-based approaches.

Oliver, et al. [31] has proposed an eigenspace model for moving object segmentation. In this method, dimensionality of the space constructed from sample images is reduced by using Principal Component Analysis (PCA). The principal component images corresponding to large eigenvalues are assumed to reflect the unchanged part of the images (background) and those corresponding to small eigenvalues correspond to a moving part (foreground). From various experiment, they claims that this method produced same accuracy with a MoG approach, but the computational run time is much faster.

Shading Model Method (SMM) [11] models the intensity at any pixel as the product of the illumination and a shading coefficient. In order to detect structural changes in a scene, the ratio of the mask pixel intensities between current frame and background frame were calculated, produced a variance values. If this value is larger than the predefined threshold, the pixel is marked as a moving object. The window mask choose to be a 3x3 pixels, since larger window will cause larger processing time because it involve a division operation in each window pixels. For updating the background, selective background updating with averaging (SBA) is implemented. Their method robustly handling large and fast changes of scene illumination, but computational time cost for calculating mask window pixel variance at each frame is an advantage.

Extended image averaging is presented in [32] and works on greyscale video imagery from a static camera. The subtraction method initializes a reference background with the new frames of the video input. Then it subtracts the intensity value of each pixel in the current image from the corresponding value in the reference background image. Any difference intensity values that exceed the threshold values were marked as a foreground pixel. The reference background images are updated by using cross intersection method between two filters. The first filter image is updated for only background pixel while the second filter image is updated for the entire pixel. Both the pixels are updated by using an Infinite Impulse Response (IIR) filter. At any pixel .if the current image is within another threshold value of the second filtered image. then the background is replaced with the second filter image values. if not the background is update then by using first filter image values.

2.2 Post Processing

The main purpose for applying the post processing module is to remove noise and unwanted foreground object from the segmented blobs. This process is crucial since noise can affect the ability of the system to identify and classify the blobs accurately. There are various factors that cause the noise in foreground detection such as:

i. Camera noise: This type of noise is caused by the camera's image acquisition components. During acquisition, fluctuating or intensity level changing of a camera are major factors that affect the amount of noise in resulting image.

ii. Background colour object noise: Certain parts of the objects may have similar colour as the reference background behind them. This resemblance causes some of the algorithms to detect the corresponding pixels as non foreground and caused the objects to be segmented inaccurately.

iii. Shadows and sudden illumination change: Shadows appears due to reflection by sun rays or light. Normally, shadow will be detected as a foreground by most of the detection algorithms. However, this problem will makes the algorithms fail to detect actual foreground objects accurately.

Post processing is defined as a process of improving the candidate foreground mask. Normally all the foreground mask produced by segmentation process contaminated with noise. These problems occur because some reason. Some of them are the system ignore any correlation between neighbouring pixels, the rate of adaptation may not match the moving speed of the foreground objects and non-stationary pixels from moving leaves or shadow cast by moving objects are easily mistaken as true foreground objects.

The first problem typically results in small false-positive (wrong pixel in the foreground) or false-negative (wrong pixel in the background) regions distributed randomly across the candidate mask. The most common approach is to combine morphological filtering and connected component grouping to eliminate these regions [12] [33] [34]. Applying morphological filtering (combination of erosion and dilation) on foreground masks eliminates isolated foreground pixels and merges nearby disconnected foreground regions. Many applications assume that all moving objects of interest must be larger than a certain size such as 30 pixels and above. Connected-component grouping can then be used to identify all connected foreground regions, and eliminates those that are too small to correspond to real moving objects.

When the background model is adapted at a slower rate than the foreground scene, large areas of false foreground commonly known as "ghosts" often occur [28]. If the background model adapts too fast, it will fail to identify the portion of a foreground object that has corrupted the background model. A simple approach to alleviate these problems is to use multiple background models running at different adaptation rates, and periodically cross-validate between different models to improve performance [24]. Sophisticated vision techniques can also be used to validate foreground detection. Computing optical flow for candidate foreground regions can eliminate ghost objects as they have no motion [24]. Colour segmentation can be used to grow foreground regions by assuming similar colour composition throughout the entire object [27].

The moving-leaves problem can be addressed by using sophisticated background modelling techniques like MoG [28] and applying morphological filtering for cleanup. For suppressing a moving shadow is much more problematic, especially for luminance-only video. Shadows cause the motion detection methods fail in segmenting only the moving objects and make the upper levels such as object classification to perform inaccurate. The proposed methods in the literature mostly use chromatic colour [24] [35] [36] or HSV colour [9] [37] information to cope with shadows and sudden light changes.

In [35], each pixel is represented by a colour model that separates brightness from the chromaticity component. A given pixel is classified into four different categories (background, shaded background or shadow, highlighted background and moving foreground object) by calculating the distortion of brightness and chromaticity between the background and the current image pixels. The approach described by Mikic et al. in [36] uses brightness and normalized red and blue colour components information to cope with shadows. Each feature is analyzed by a posterior probability estimator that computes membership probabilities for the three classes (background, foreground, and shadow). The algorithm literates between the three estimators by using the output of one as the prior probability input to the next. The method presented in [24] adopts a shadow detection scheme which depends on two heuristics which are chromatic colour value and lightness variable for each pixels. Shadow is detected and marked if distance chromatic colour measure and ratio of lightness between current and reference is greater than threshold value.

An attempt to use HSV colour space for shadow detection is presented in [9] [37]. They claim that HSV colour space correspond closely to the human perception of colour and more accurate in distinguishing shadows than the RGB space. During implementation, the pixel is marked as a shadow if ratio of V, difference of S and absolute difference of H for current frame and background frame within the range of the define threshold values. However the choice of threshold parameters is less straightforward, and normally is done empirically with the assumption that the chrominance of shadowed and non-shadowed points even if could vary, does not vary too much.

2.3 Feature Extraction

It is very important to recognize the type of a detected object in order to track it reliably and analyze its activities correctly. Before object can be recognized, an important feature need to be extract from it, usually features are selected based on its' ability to distinguish or separate the object classes with a minimum overlapping between it. Currently, there are two major feature extraction approaches for moving object classification which are shape-based and motion-based methods [2]. Shape-based methods make use of the objects' two dimension spatial information whereas motion-based methods use temporally tracked features of objects for the classification solution. This thesis focuses on a method to recognize object based on its shape features because human and vehicle can easily discriminate based on it shape. Therefore, the remainder of the section will concentrate on this features which can be used in most applications.

Shape is an important visual feature and it is one of the primitive features for image content description. In the recent years, shape research has been driven mainly by object recognition. As a result, techniques of shape representation and description mostly target at particular applications. Effectiveness or accuracy is the main concern of these techniques. The variety of shape representation techniques can be classified into contour based and region based method. Under these classes, the different methods are further distinguished between structural and global based on whether the shape is represented as a whole or represented by sub parts (primitives).

The whole hierarchy of the feature extraction based on shape is shown in Figure 2.2.

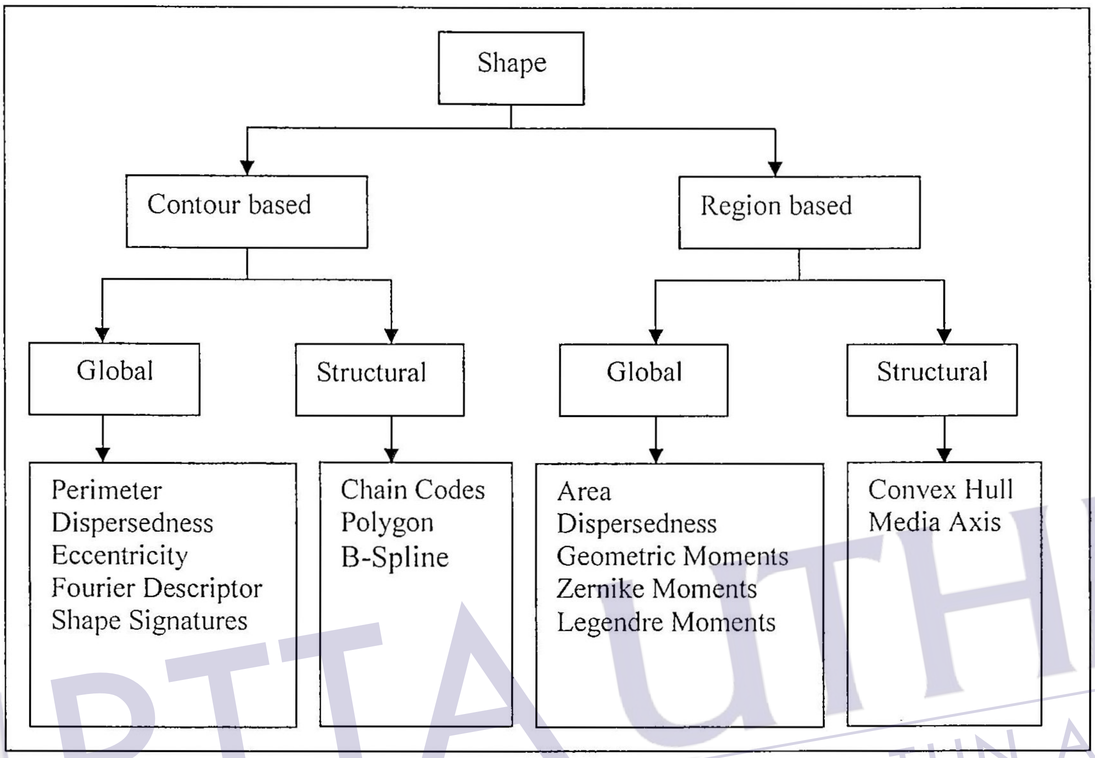

Figure 2.2: Hierarchy of shape based representation and description techniques.

Region and contour shape techniques exploit shape region or boundary information. There are generally two types of very different approaches for the shape modelling: global approach and structural approach. Global approaches do not divide shape into sub parts. A feature vector derived from the boundary is used to describe the shape. The measure of shape similarity is either point-based matching or feature-based matching. Structural approaches break the shape boundary into segments, called primitives using a particular criterion. The final representation is usually a string or a graph, the similarity measure is done by string matching or graph matching.

However structural approaches have some drawbacks, because there is no formal definition for an object or shape, the number of primitives required for each shape is unknown [38]. Therefore, the success of applying this method depends on the a priori knowledge of the shape boundary features during learning. So it is impractical to use it for general applications, because it is impossible to know in advance the types of objects primitives in a real time environment. Furthermore, structural approach fails to capture global shape features which are equally important for the shape representation. Because of structural representation do not represent topological structure of the object. Therefore it is not suitable to be implemented in the system. In this case, global features are more reliable. The following subsection will discuss some of the well known global contour and region based shape representation techniques.

2.3.1 Global Contour Based Shape Representation Techniques.

Global contour shape representation techniques usually compute a multi-dimensional numeric feature vector from shape boundary information. Matching between shapes is a straightforward process, which is usually conducted by using a metric distance, such as Euclidean distance or city block distance. Point or point feature based matching is also used in particular applications.

Common simple global descriptors like dispersedness (perimeter^2/area), eccentricity (length of major axis/length of minor axis) and major axis orientation have been implemented in various of surveillance system [12][39][40].These simple global descriptors usually can only discriminate shapes with large dissimilarities. Therefore,

REFERENCES

1. Radke, RJ , Andra S, Al-Kofahi O, and Roysam B. Image Change Detection Algorithms: A Systematic Survey. *IEEE Transactions on Image Processing* 2005; 14(3): 294-307.

2. Omar Javed, Saad Ali, and Mubarak Shah. Online detection and classification of moving objects using progressively improving detectors. *IEEE Computer Society Conference on Computer Vision and Pattern Recognition 2005.* San Diego USA.

3. Wang L, Hu W, and Tan T. Recent developments in human motion analysis. *Pattern Recognition* 2003; 36(3): 585-601.

4. Thomas M and Ngan, KN. Automatic segmentation of moving objects for video object plane generation. *IEEE Trans. Circuits and System for Video Technology* 1998; 8(5): 525-538.

5. Jolly, MPD, Lakshmanan S, and Jain, AK. Vehicle segmentation and classification using deformable templates. *IEEE Trans. Pattern Analysis and Machine Intelligence* 1996; 18(3):293-308.

6. Cheung, SC and Kamath C. Robust techniques for background subtraction in urban traffic video. *Proceedings of Electronic Imaging: Visual Communications and Image Processing* 2004 .San Jose California.

7. Lee B and Hedley M. Background Estimation for Video Surveillance. *Image and Vision Computing* 2002. New Zealand ; 315-320.

8. Foresti, GL, Micheloni C, Snidaro L. Advanced visual-based traffic monitoring systems for increasing safety in road transportation. *Advances in Transportation Studies an International Journal* 2003; 1(1):22-47.

9. Cucchira R, Piccardi M, and Prati A. Detecting moving objects, ghost and shadow in video stream. *IEEE Trans. On Pattern Analysis and Machine Intelligence* 2003: 25: 1337-1342.

10. Heikkilä M and Pietikäinen M. A texture-based method for modeling the background and detecting moving objects. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 2006; 28(4): 657-662.

11. Zeljkovic V, Trpovski Z, and Senk V. Improved illumination independent moving object detection in real world video sequences. *In Pro. 4th EURASIP Conf. focus on Video/Image Processing and Multimedia Communication* 2003: 2.

12. Brown, LM. View Independent Vehicle/Person Classification. *ACM 2nd Int'l Workshop on Video Surveillance & Sensor Networks* 2004.Columbia Universit.New York City.

13. Collins , RT. A system for video surveillance and monitoring: VSAM final report. *Technical report CMU-RI-TR-00-12* May 2000; Robotics Institute, Carnegie Mellon University.

14. Horn, BKP, and Schunck, BG. Determining Optical Flow: a retrospective. *Artificial Intelligence* 1993; 59: 81-87.

15. Lucas, BD.,1984. Generalized Image Matching by the Method of Differences, PhD Dissertation, Carnegie Mellon University, USA.

16. Anandan P. A Computational Framework and an Algorithm for the Measurement of Visual Motion, *International Journal on Computer Vision* 1989; 2: 283-310.

17. Nagel, HH., On a Constraint Equation for the Estimation of Displacement Rates in image Sequances, *IEEE Transaction on Pattern Recognition and Machine Learning* 1989; 11:13-30.

18. Uras S, Girosi F, Verri A, and Torre V, A Computational Approach to Motion Perception. *Bioogical. Cybernatic* 1988; 60:79-97.

19. Camus T. Real-time quantized optical flow, *The Journal of Real-Time Imaging* 1997; 3: 81-87.

20. Proesman M, Val Gool L, Pauwels E, and Oosterlinck. Determination of optical flow and its discontinuities using non-linear diffusion. *In 3rd Euorepean Conference on Computer Vision* 1994. Stockholm, Sweden; 2: 295-304.

21. Barron J, Fleet D, Beauchemin S, and Burkit T. Performance of optical flow techniques. *Proc. of the IEEE on Computer Vision and Pattern Recognition* 1992; 236-242.

22. Galvin B, McCane B, Novins K, Mason D, and Mills S. Recovering Motion Fields: An Evaluation of Eight Optical Flow Algorithms. *Proceedings of Ninth British Machine Vision Conference* 1998. Southhampton. United Kingdom.;195-204.

23. Neri A, Colonnese S, Russo G, and Talone P. Automatic moving object and background separation. *Signal Processing* 1998; 66(2):219-232.

24. Elgammal A, Harwood D, and Davis L. Non-parametric model for background subtraction. *In Proc. Of IEEE International Conference on Computer Vision 99 Frame-rate workshop* 1999. Dublin Ireland.

25. Karman, KP and Von Brandt A. Moving object recognition using an adaptive background memory. *Time-Varying Image Processing and Moving Object Recognition* 1990. Elsevier Science Publishers B.V.

26. Dr Alan M. McIvor. Background Subtraction Techniques. *Image and Visoin Computing New Zealand* 2000. Hamilton.New Zealand.

27. Toyama K, Krumm J, Brumitt B, and Meyers B. Wallflower: Principles and practice of background maintenance. *International Conference on Computer Vision* 1999. Kerkyra, Greece; 255-261.

28. Stauffer C and Grimson, WEL. Learning pattern of activity using real-time tracking. *IEEE Trans. Pattern Analysis and Machine Intelligence* 2000; 22(8): 747-757.

29. Li D. Moving object detection by block comparison. *Proc. 7th int. Conf. on Electronics Circuits and Systems* 2000. Jounieh, Lebanon;1: 341–344.

30. Yang, YH and Levine, MD. The background primal sketch: An approach for tracking moving objects. *Machine Vision Applicat.* 1992; 5:17–34.

31. Oliver, NM, Rosario B, and Pentland, AP. A Bayesian computer vision system for modeling human interaction. *IEEE Trans. PAMI* 2000; 22(8): 831-843.

32. Sacchi C, Gera G, Marcenaro L, and Regazzoni, CS. Advanced image-processing tools for counting people in tourist site-monitoring applications. *Signal Processing* 2001; 81(5): 1017-1040.

33. Yang T, Li, SZ , Pan Q, and Li J. Real-time and accurate segmentation of moving objects in dynamic scene. *Proceedings of the ACM 2nd international Workshop on Video Surveillance &Amp; Sensor Networks* 2004. New York. USA. October 15 – 15: 136-143.

34. Kang S, Palk J, Koschan A, Abidi B, and Abidi, MA. Real-time video tracking using PTZ cameras. *Proc. of SPIE 6th international conference on Quality Control by Artificial Vision* May 2003.Gatlinburg.TN; 5132: 103-111.

35. Thongkamwitoon T, Aramvith S, and Chalidabongse, TH. An Adaptive Real-time Background Subtraction and Moving Shadows Detection. *Proceeding of International Conference on Multimedia and Expo* 2004. Taipei. Taiwan..

36. Mikic I, Cosman P, Kogut G, and Trivedi, MM . Moving Shadow and Object Detection in Traffic Scenes. *Proc. Int'l Conf. Pattern Recognition* Sept. 2000. Washington, DC;1: 321-324.

37. Cucchiara R, Grana C, Piccardi M, Prati A, and Sirotti S. Improving Shadow Suppresion in Moving Object Detection with HSV Color Information. *Poc. IEEE Intelligent Transportation Syatem Conference* August 25-29 2001. Oakland(CA). USA.

38. Dengsheng Zhang, 2002. Image Retrieval Based on Shape, Ph.D Thesis, Monash University.

39. Lipton, AJ, Fujiyoshi H , and , Patil, RS. Moving Target Classification and Tracking from Real-time Video. *In Proceedings of the 4th IEEE Workshop on Applications of Computer Vision* 1998.IEEE Computer Society. Washington DC.

40. Zhou Q and Aggarwal J. Tracking and Classifying Moving Objects from Video. *In Proc. 2nd IEEE Int'l Workshop on Performance Evaluation of Tracking in Surveillance* Dec 9 2001. Kauai, Hawaii.

41. Daniel Toth and Til Aach. Detection and Recognition of Moving Objects Using Statistical Motion Detection and Fourier Descriptors. *12th International Conference on Image Analysis and Processing* 2003. Washington, DC.

42. Kim C and Hwang, JN. Object-based video abstraction for video surveillance systems. *IEEE Transactions on Circuits and Systems for Video Technology* 2002; 12(12): 577-586.

43. Squire, DM and Caelli, TM. Invariance Signature : Characterizing Contours by their Depatures from Invariance. *Computer Vision and Image Understanding* 2000; 77:284-316.

44. Hu, MK. Visual Pattern Recognition by Moment Invariants. *IEEE Transaction on Information Theory* 1962; 8(2): 179-187.

45. Yea Hong Yoon and Kang Hyun Jo. Hands Shape Recognition using Moment Invariant for the Korean Sign Language Recognition. *The 7th Korea-Russia International Symposium on Science and Technology* 2003. South Korea.

46. Lou T, Kramer K, Goldgof D, Hall L, Sampson S, Remsen A, Hopkins T. Learning to recognize plankton. *IEEE International Conference on Systems, Man & Cybernetics* October 2003. Washington, D.C; 888-893.

47. Hsu, CJ, Chen, CF, Lee C, and Huang, SM. Airport Pavement Distress Image Classification Using Moment Invariant Neural Network. *22th Asian Conference on Remote Sensing* 2001. Singapore; 216-220.

48. Simon Xinmeng Liao, 1993. Image Analysis by Moments. Ph.D Thesis. University of Manitoba, Canada.

49. Prokop, RJ and Reeves, AP. A survey of moment-based techniques for unoccluded object representation and recognition. *CVGIP: Graphical Models and Image Processing* 1992; 54:438-460.

50. Burges, CJC. A tutorial on support vector machines for pattern recognition. *Data Mining Knowledge Discovery* 1998; 29(2):121–167.

51. Osuna E, Freund R, and Girosi F. Training support vector machines:An application to face detection. *Proc. Computer Vision and Pattern Recognition* 1997; Puerto Rico:130–136.

52. Lou T, Kramer K, Goldgof D, Hall L, Sampson S, Remsen A, Hopkins T. Learning to recognize plankton. *IEEE International Conference on Systems, Man & Cybernetics* October 2003; Washington, D.C: 888-893.

53. Dumais, ST, Platt J, Heckerman D, and Sahami M. Inductive learning algorithms and representations for text categorization. *Proc. ACM-Conf. Information and Knowledge Management (CIKM98)* Nov 1998. Bethesda, Maryland; 148–155.

54. Gupte S, Masoud O, Papanikolopoulos, NP. Detection and classification of vehicles. *IEEE Trans. on Intelligent Transportation Systems* 2002; 3(1): 37-47.

55. Gonzalez, RC, Woods, RE. *Digital Image Processing*, 2nd ed.; Prentice Hall: New Jersey, 2002.

56. Mcandrew A. *Introduction to Digital Image Processing with Matlab*, 1st ed.;Thomson Course Technology:United State of America, 2004.

57. Zhang H, Guan J, and Sun GC. Artificial neural network-based image pattern recognition. *In Proceedings of the 30th Annual Southeast Regional Conference* 1992;ACM Press. New York.

58. Rifkin, RM.,2002. Everything Old is New Again : A Fresh Look at Historical Approaches in Machine Learning, Ph.D Thesis, Massachusetts Institute of Technology, Boston USA.

59. W. G. Lin and Sh. Sh. Wang, A note on the calculation of moments, Pattern Recognition Letters, 15:1065-1070, 1974.

60. http://cbcl.mit.edu/software-datasets/