University of Windsor

Scholarship at UWindsor

Electronic Theses and Dissertations

Theses, Dissertations, and Major Papers

1-1-2006

Background modeling for intelligent video surveillance system.

Simeon Immanuel Kiran Indupalli
University of Windsor

BACKGROUND MODELING FOR INTELLIGENT VIDEO SURVEILLANCE SYSTEM

by

Simeon Immanuel Kiran Indupalli

A Thesis
Submitted to the Faculty of Graduate Studies and Research
through the Computer Science
in Partial Fulfillment of the Requirements for the Degree of Master of Science
at the University of Windsor

Windsor, Ontario, Canada
2006

Library and
Archives Canada

Published Heritage
Branch

395 Wellington Street
Ottawa ON K1A 0N4
Canada

Bibliothèque et
Archives Canada

Direction du
Patrimoine de l'édition

395, rue Wellington
Ottawa ON K1A 0N4
Canada

Your file Votre référence
ISBN: 978-0-494-35961-7
Our file Notre référence
ISBN: 978-0-494-35961-7

Abstract

Despite the dramatic growth of digital image and video in recent years, many challenges remain in enabling computers to interpret visual content. This thesis addresses the initial stages of building an intelligent video surveillance system. The central contribution of this work is developing a framework which models the scene background and segmenting the moving foreground objects in video image data. This work is divided into two sections.

In the first section, a background model is being designed dynamically with no prior knowledge of the scene. A simple histogram based method is used to accomplish this task.

The second step involves segmenting the moving foreground objects using a clustering mechanism. We have used a simple K-means clustering, where the value of K is two. We have implemented our methods in HSV color space. The results obtained were quite satisfactory in different real environments.

Dedication

To my parents

Acknowledgements

"The fear of the LORD is the beginning of knowledge" Proverbs 1:7

With a grateful heart I thank Him for all the blessings and knowledge in pursue of this worldly wisdom.

Secondly, I would like to express my sincere gratitude to my advisor Dr. Boubakeur Boufama. Without his support this thesis work would be impossible in many ways. His continuous support and insightful guidance helped in every stage of my master's degree. I feel privileged to have him as my advisor.

I would like to thank the Networks Centres of Excellence Auto21 for funding this thesis. I thank Dr. Imran Ahmed and Dr. Jonathan Wu for serving on my thesis advisory committee and providing me with valuable feedback.

I gratefully acknowledge my colleague Ahsan Ali for his helpful discussions and suggestions. I render special thanks to my friend Mayuran: he has been with me right from the first day of my arrival here on campus. I acknowledge his help and support.

Finally, many thanks to all my friends on and off the campus; it is hard to imagine a life in Windsor without them.

Contents

List of Tables

List of Figures

Chapter 1

Introduction and Motivation

1.1 Introduction

This chapter gives an introduction to the origin of computer vision, and related fields associated with it. In the subsequent sections, we have given a brief narration on the concept of intelligent video surveillance systems. The motivation and contributions follows this section. A detailed summary on the structure of this thesis was given at the end.

1.2 Problem of Visual Representation

"....Consider the name of signs the mind makes use of for the understanding of things, or conveying its knowledge to others. For, since the things the mind contemplates are none of them, besides itself, present to the understanding, it is necessary that something else, as a sign or representation of the thing it considers, should be present to it."
-Jon Locke
Essay Concerning Human Understanding - 1690

Human brains respond alike when they see a known object. For instance, if two people see a cat they respond alike saying it is a cat. What might be the underlying factor behind this analogy? This question leads us to the philosophy of mind (Cummins, 1989). Many applications are possible if we can interpret this analogy to machines, but many implications follow this delegation of human interpretation to machines.

Computer Vision is the enterprise of automating and integrating a wide range of processes and representations used for vision perception. The field of Computer Vision partially collide with other research areas like *image processing* (transforming, encoding and transmitting images) and *statistical pattern classification* (statistical decision theory applied to general patterns, visual or otherwise).

Visual perception is the relation of visual input to previously existing models of the world. There is a large representational gap between the image and the models ('image', 'concepts') which explain and describe the image information. To bridge the gap, computer vision systems have a range of representations connecting the input and the 'output' (a final description, decision, or interpretation). Computer vision then involves the design of these intermediate representations and the implementation of the algorithms to construct them and relate them to one another [52].

1.3 The Origin

The field of Computer Vision was not very famous until the late 1970's. The actual work started when computers could manage the processing of large data sets such as images. However, many methods and applications are still in the state of basic research, but many methods find their way into commercial products, where they constitute a small part of a larger system and can solve the complex tasks [51].

As mentioned earlier, computer vision collaborates with different other fields of science. In physics it plays a significant role in understanding the processes in which electromagnetic radiation, (typically in the visible or the infra-red range and the surface reflectance of the objects) are measured by the image sensor to produce image data. In the area of artificial intelligence, Computer Vision deals with robot navigation through some environment. This type of processing needs input data provided by a computer vision system. Many more scientific domains including signal processing, neurobiology, statistics, optimization and geometry need computer vision systems in their respective applications.

With all of the demand in various application domains as mentioned above, computer vision systems have been introduced in the field of video surveillance applications. As defined in [36], "It is the science and technology of machines that see". This technology provides the background to develop intelligent systems based on the information obtained from the images. The applicability of vision systems in video surveillance are given below:

- Detecting events for visual security surveillance.
- Modeling objects or environments for industrial inspection or topographical modeling.

And many more real-time applications are possible with the robust expansion of this field.

1.4 Intelligent Video Surveillance Systems

Someone approaches the door of a national security services facility. A security camera watching this person does more than simply sending this information to a guard. It scans the person's face and runs this video through facial recognition program that checks against a criminal watch list and get list. The person tries to break in, the perimeter breach alarm sets off and alerts the control room. It locks all the entrance doors. Text messages will be sent to all the personnel across the facility indicating the possible threat. The cameras tracking this person would now initiate an activity recognition program to identify possible suspicious actions and use of weapons possessed by this person.

The above scene imitates the scenario in a science fiction story. It is not surprising to say, that we will have this systems in action very soon. Many organizations prefer video surveillance systems doing more than just capturing the videos.

The intelligent video surveillance systems improve the security standards and solve the scalability issues arising due to the traditional video surveillance systems. Intelligent system will enable its users to understand as much as 10 times more because they can just look at mug shots instead of raw footage.

1.5 Motivation

Keeping the versatility of the above system in mind, a number of issues have to be taken care of, to model such system. Background subtraction is at the heart of intelligence video surveillance applications. To perform this operation, one has to provide a background image that has no foreground objects. However, in real-time conditions it is not always possible to capture the background in advance. In that situation dynamic background modeling plays an important role. In other words, dynamic background modeling is a desired starting step for any robust video surveillance system.

Segmenting the foreground objects follows the background modeling. In order to segment a foreground object, one has to provide a proper threshold value based on which segmentation shall be performed, but it needs manual observation. Automatic threshold calculation will considerably eliminate the human intervention at the middle of the process.

1.6 Issues addressed in this thesis and Contributions

This thesis mainly focuses on modeling the background in busy areas and segmenting the foreground moving objects. We have modeled the background using a histogram based method and a segmentation technique was also proposed to segment the moving foreground objects from the background. We have used k-means clustering technique in our segmentation algorithm. We did an informal comparison on various color spaces in implementation phase. We chose HSV color space, which has an edge over other color spaces in regard to shadows and noise.

1.7 Thesis Structure

The rest of the thesis is organized as follows:

In chapter 2, we have given a brief introduction to video surveillance systems followed by background modeling and segmentation issues. Chapter 3 gives an overview of the methods proposed by different researchers in the literature. Chapter 4 is divided into two sections: section one deals with background modeling, the other section explains the detailed process of segmentation for segmenting the foreground objects. Results of the proposed methods were presented in Chapter 5. We conclude this thesis with some remarks and future work in Chapter 6.

1.8 Conclusion

This is an introductory chapter on various issues relating to video surveillance and computer vision. We briefly mentioned the topic of intelligent video surveillance with a narrative description. Our motivation behind this work was presented in the motivation section.

Chapter 2

Intelligent Video Surveillance System: An Overview

This chapter focuses on the overview of an intelligent video surveillance system. We have given a detailed description of the functional architecture of an intelligent video surveillance system. In the subsequent sections, a detailed elaboration of issues relating to background modeling and segmentation were given.

2.1 Overview of Intelligent Video Surveillance System

With recent advances in computer technology visual surveillance has become one of the popular areas for research and development. In the earlier days, visual surveillance involved manual observation of video sequences for several hours. From the perspective of real-time threat detection, it is a well known fact that human visual attention drops below the acceptable level even when trained personnel are assigned to the task of visual monitoring [41].

As a whole, visual surveillance systems seek to automatically identify events of interest in a variety of situations [30]. Surveillance cameras are installed in many public areas to improve safety, and computer-based image processing is a promising means to handle the vast amount of image data generated by large networks of cameras. The task of an integrated surveillance system is to warn an operator when it detects events which may require human intervention, for example to avoid possible accidents or vandalism. A lot of promising applications are based on successful visual surveillance systems, such as vehicle guidance [24], object tracking for security surveillance [25], traffic monitoring, detection of abnormal activity and threat evaluation [44]. The general framework of these applications groups the number of different computer vision tasks such as detection, tracking and classification of objects of interest in image sequences, and understands and describes the activities involving the objects.

2.1.1 Introduction

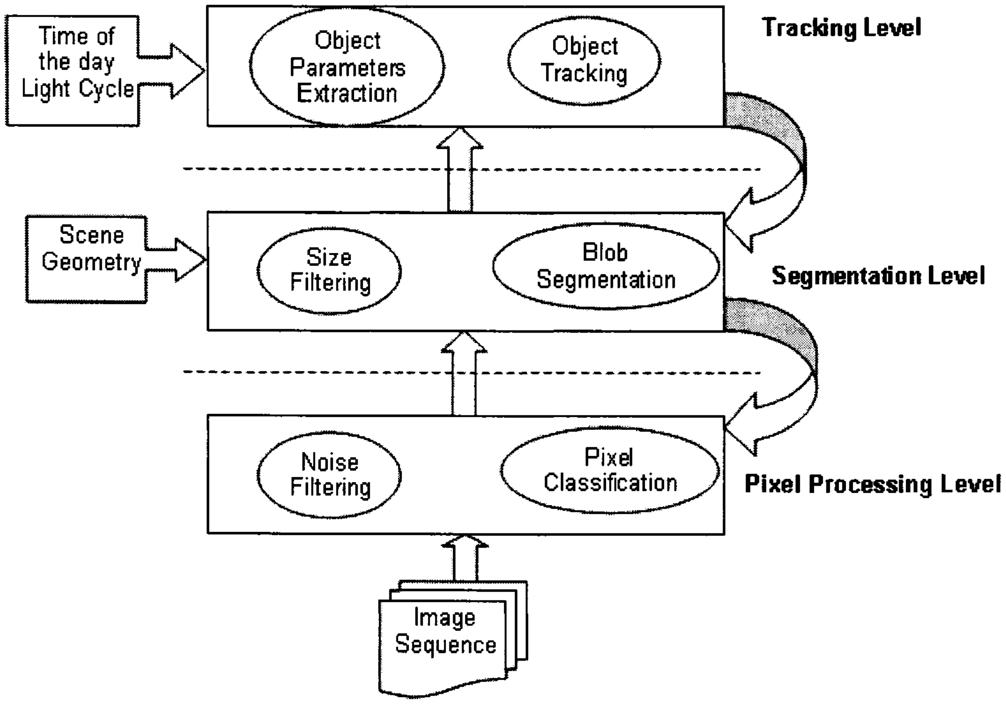

Figure 2.1: Building blocks of intelligent video surveillance system [42]

In Fig. 2.1, an abstract overview of an intelligent video surveillance system has been given. The process starts from the collection of video frames sequentially from an uncalibrated camera. These video frames undergo further processing, for analysis and information retrieval. The whole process has been divided into the following functional blocks:

1. Pixel processing level
2. Object segmentation level
3. Tracking level

This thesis mainly focused on the first building block of the system. Wece have assumed to model a system which can dynamically model the background from a busy scene and segment the moving foreground objects.

2.1.2 Pixel processing level

Pixel classification is a basic step for any intelligent video surveillance system. The accuracy and versatility of the whole application depends on this step. Different techniques are being used to accurately classify the pixels, based on the intensity, color and pixel position [34], [8], [25].

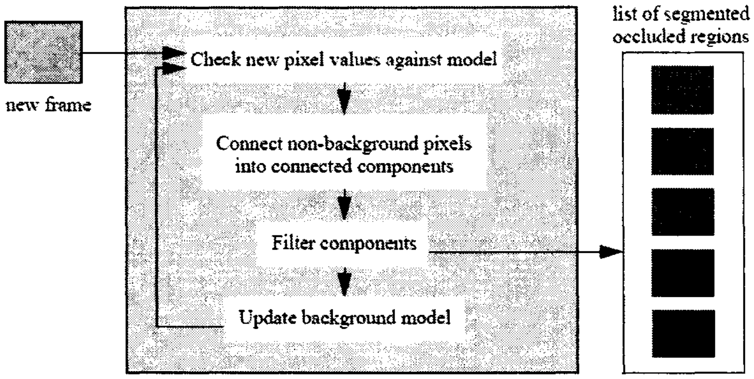
new frame
Check new pixel values against model
Connect non-background pixels into connected components
Filter components
Update background model
list of segmented occluded regions

Figure 2.2: A Scenario at pixel processing level [45]

The pixel processing can be performed at three different levels of abstraction, depending on the type of application and processing speed in real-time.

Background Subtraction

The most popular method is to classify pixels into foreground and background categories. It is a simple method by which incoming video frames shall be subtracted from a background image. The sample result of the background subtraction can be seen Fig 2.3.

However, one should know the background scene before subtracting the incoming image frame. The background image can be attained in two ways, one by capturing the empty scene with no moving objects prior to the starting of the process. Dynamic background modeling is another approach, which models the background dynamically with no prior knowledge of the scene.

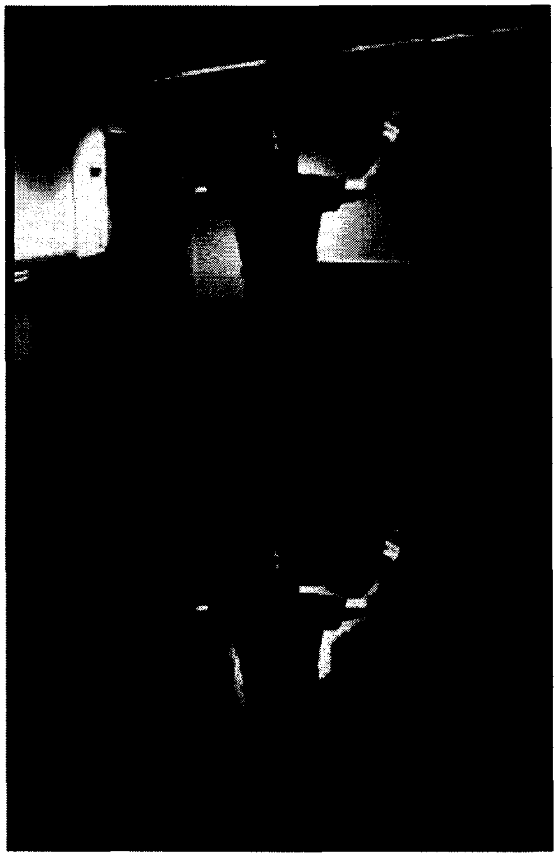

Figure 2.3: A person identified using background subtraction

The dynamic background modeling capable to model the approximate background using various techniques. The Gaussian Mixture model, linear predictive filter, Kalman filter were among the famous contributions in this area. In this thesis, we have proposed a histogram based model to model the background. The proposed model make use of some statistical information captured from the image sequence over a period of time. The detailed explanation can be found in the later chapters.

Limitations of Background Subtraction

Background subtraction heavily depends on the background model of the scene. A constant updating of the background model is necessary for better segmentation* in later stages.

Frame Differencing

Frame differencing is another technique to identify the moving foreground objects. The difference of two consecutive frames is further processed to identify the foreground regions. However, this method has some severe drawbacks in accurate segmentation of the moving objects. The result of such operation is shown Fig. 2.3.

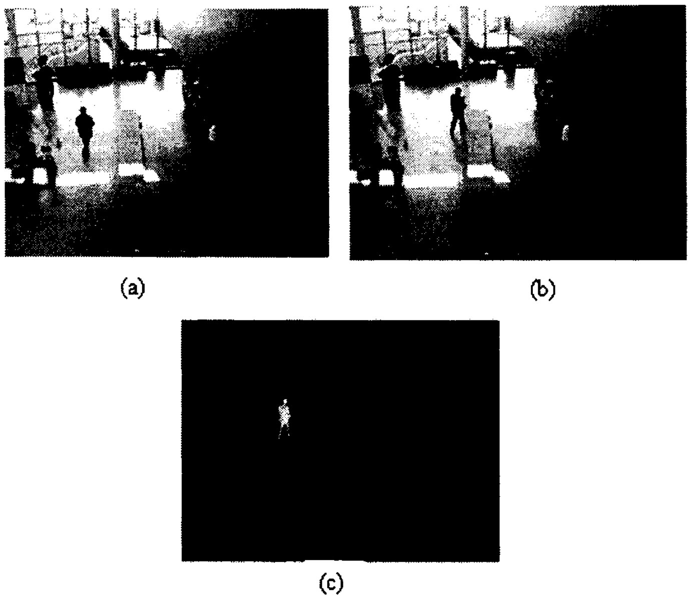

Figure 2.4: An example of frame difference technique: (a) and (b) are two consecutive frames (c) is the resulting image obtained using frame differencing.

Limitations of Frame Differencing

One major problem with frame differencing technique is it leaves holes in the resulting image. This is quite problematic in later stages of the application, especially when updating the model of the background as this error linger for a longers period of time.

Optical Flow

Optical flow is a concept for estimating the motion of the objects within a visual representation. Typically the motion is represented as vectors originating or terminating at pixels in a digital image sequence [36].

Limitations of Optical Flow methods

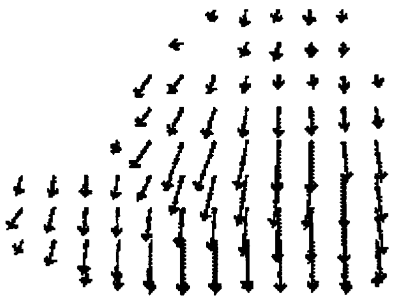

Figure 2.5: An example of optical flow: Moving objects identified using these motion vectors

Once the optical flow is estimated from the consecutive image frames the segmentation is implied based on the flow of motion. Optical flow techniques are considerably slow due to the complexity and computational time.

2.1.3 Segmentation level

Segmentation is the next level after the pixel process. It deals with the identification of the moving region in the image frame. The object segmentation has different problems to deal with ranging from noise, object shape and size of the object. For segmenting the objects people have used different techniques ranging from threshold calculation [15], Gaussian mixture models [5],[25], Hidden Marcov Models etc.

Segmentation suffers from different problems arising due to different conditions. The accuracy of tracking is very much depending on the perfect segmentation of the foreground object. The following factors cause problems in segmentation:

- *Foreground aperture* - the foreground color has a close match with the background color.
- *Choosing threshold value* - we need to choose possible threshold value in order to separate the actual foreground pixels, which need some observation on pixel intensities.

This thesis also contributes a method addressing the above mentioned problems. The detailed discussion on this topic can be found in later chapters.

2.1.4 Tracking level

Tracking is the final stage of the video surveillance system. At the tracking level the movements of the objects are tracked and suspicious behavior of the objects are closely observed. Various factors are taken into consideration such as object motion, geometry etc. The results of tracking can be used in different real-time application.

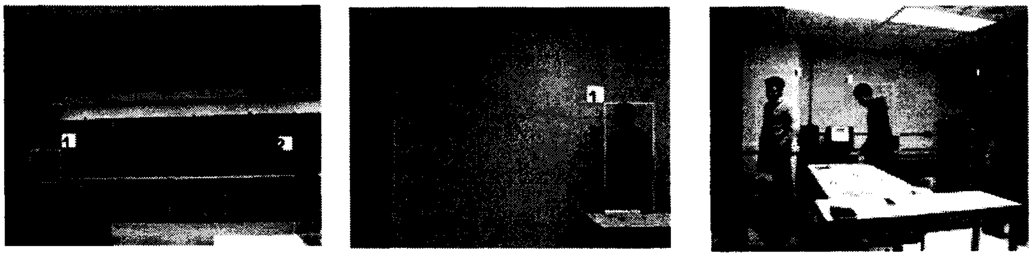

Figure 2.6: Some tracking results taken from [35]

The detailed implementation and applications of various tracking algorithms can be found in the thesis work of Ali et al. [35].

2.2 Background Modeling and Segmentation: An overview

In all the video surveillance applications, the effective way for segmenting foreground objects is to suppress the background points. To achieve this goal an accurate and adaptive background model is often desirable.

In this section we have addressed various stages in background modeling and problems in real-time. We have also given a brief explanation of different problems while segmenting the foreground objects.

2.2.1 Characterizing the issues in Background Modeling

The background modeling process has been characterized in three different sections:

Background Model Initialization

This is the first step in background modeling algorithms. The initial model is obtained by a short training period with no foreground objects when they are present in the scene [7].

Background Model Generation

This step gives the estimated background from a short training period. In other words, it dynamically models the background with moving objects present in the scene [43].

Background Model Maintenance

This is a difficult step in background modeling. In real time situations the observed scene undergoes different changes. The background model has to be updated accordingly with the changes in the scene.

2.2.2 Background modeling in real time

Background usually contains nonliving objects; those remain passive in the scene. The background objects may be stationary, such as walls, doors and room furniture, or they may be non stationary objects like tree branches, wavering bushes and moving escalators. These background objects often undergo various changes in a course of time due to the changes in brightness caused by changing weather conditions or the switching off the lights.

With respective to the above reasons, background image can be described as a combination of static and dynamic pixels. The static pixels belong to the stationary objects, and the dynamic pixels are associated with non-stationary objects. Below are the examples of various scene characteristics:

Time of the day

Time of the day comes under updation of the background scene. This is a common problem while updating the background when the light gradually fades out.

Figure 2.7: Gradual lighting change during the day

Sudden illumination changes

Quick illumination changes completely alter the color characteristics of the background, thus increases the intensity of background pixels. The result of this change will cause a false detection of background as foreground, The worst cases, the whole image might appear as foreground.

Moving branches at the background

This is a very common scenario in almost all the outdoor video surveillance systems. The moving branches of the trees in the background will cause problems to accurately model the background.

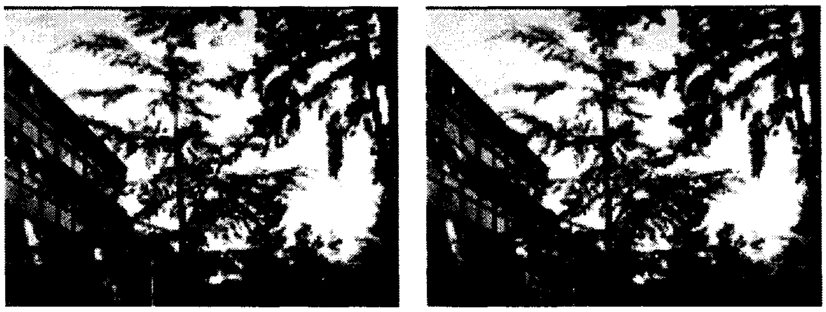

Figure 2.8: Moving branches of the trees at the background

Uncontrolled moving objects in the background

It is hard to identify the actual background scene if objects in the background move continuously. Many algorithms use a scene without moving objects, but this kind of assumption will put some serious limitations in highly densed areas.

Figure 2.9: Continuous moving objects in the background

Shadows

Shadows create problems while modeling the background. Shadows cast by the objects are classified as foreground, due to false illumination in the shadow region.

2.2.3 Functional expectations of Background Modeling

To develop a general background scene, a background model must be able to satisfy the following conditions:

1. Should represent the appearance of a static background pixel
2. Should represent the appearance of a dynamic background pixel
3. Self-evolve to gradual background changes
4. Self-evolve to sudden once-off background changes

Background is usually represented by image features at each pixel. The features extracted from an image sequence can be classified into three types: spectral, spatial, and temporal features. Spectral features can be associated with gray-scale or color information. Spatial features can be associated with gradient or a local structure, and temporal features can be associated with inter frame changes at the pixel. Many existing methods utilize spectral features (distributions of intensities or colors at each pixel) to model the background [4],[5],[7],[9].. In order to be robust to illumination changes, some spatial features are also exploited [2],[10],[12]. The spectral and spatial features are suitable to describe the appearance of static background pixels.

Recently, few methods have introduced temporal features to describe the dynamic background pixels associated with non stationary objects [6],[13],[14]. There is, however, a lack of systematic approaches to incorporate all three types of features into a representation of a complex background containing both stationary and non stationary objects. The features that characterize stationary and dynamic background objects should be different. If a background model can describe a general background, it should be able to learn the significant features of the background at each pixel and provide the information for foreground and background classification.

2.2.4 Segmentation of moving objects

Video object segmentation is a challenging step. Segmentation is nothing but figuring what is background and what is foreground. Segmentation can be done on variety of factors like intensity, color, size and location.

Problems in segmentation

Segmentation stage contributes a lot while tracking the moving objects. The accurate segmentation of foreground objects is difficult due to shadows in the object region, and foreground aperture.

Shadows in the foreground region

Shadows create problems during segmentation, shadows are categorized into cast shadows and self shadows. Cast shadows are caused by the moving objects on the side of the object, where as self shadows are cast by the object on the object itself (shadows cast by the folding of a shirt).

In [50], Prati et al conducted a comprehensive survey on the shadow detection of moving objects. A study was conducted on a deterministic character, shadows are eliminated based on the assumption that the chromaticity of the shadow region is slightly dark compared to actual object region.

Foreground aperture is another problem occurs while segmentation, when the foreground object color matches close the background object, part of the foreground object disappears after segmentation. A better segmentation mechanism can identify these problems.

2.3 Conclusion

This chapter is a introductory chapter on intelligent video surveillance applications. The filed of intelligent video surveillance, gaining popularity in the recent years, many interesting methodologies being proposed by different researchers for real time applications. Freshers to this field can benefit from this chapter, as they come across various problems in modeling the background and segmentation.

Chapter 3

Literature Review

Background modeling is at the heart of any background subtraction algorithm. These algorithms have been classified into two broad categories, they are non-recursive and recursive [1].

3.1 Non-recursive Techniques

In Non-recursive techniques, the algorithms have to process a buffer of previous N video frames and estimate the background image based on the temporal variation of each pixel within the buffer. These Non-recursive algorithms are highly adaptive and they do not depend on the history beyond the stored buffer. However, it is expensive to model a scene with slow moving objects using these algorithms. Following are the major algorithms, that fall under this category:

3.1.1 Median filter

A median based method was proposed in [6]and [19]. A slightly modified version was proposed by Howe et al in [3], this method recursively updates the background based on the previous three image frames. With a similar idea, Long et al proposed an adaptive smoothness algorithm, which finds the intervals of stable intensity and uses the heuristics to choose the longest interval [18]. Similar to this proposed approach, Gutchess et al. in [7], proposed a method called the local image flow algorithm. This method [7] also depends on the intervals of stable intensity of an image sequence. Along with observing the intervals of stable intensity, the vicinity of neighboring pixels are also taken into consideration for estimating the background. This measure eliminates the blending problem, when foreground objects stay in the image for a longer period. An optical flow technique is being implemented, by summing the distances of each vector head in the local neighborhood using the Gaussian mixture. The main strength of this algorithm in [7], is to take a decision, whether the neighboring pixel belongs to either a background or a foreground.

3.1.2 Linear predictive filter

Toyama et al. proposed an approach which uses wiener filter[28], for background modeling the background. Addressing various issues in background modeling, they have proposed an approach at pixel, region and frame level to construct a robust background model which can deal with most of the real time problems. At pixel level if the prediction deviates far from the expected value, the pixel shall be considered as foreground. For a given pixel, the linear prediction of its next value is given by

$$S_t = \sum_{K=1}^{P} a_k s_{t-k} \tag{3.1}$$

This filter uses the past P values to predict pixel S_t at time t, s_{t-k} is the past value of a pixel and a_k is the prediction coefficient. At region level, they consider to segment the whole object rather than some isolated pixels. At frame level,a representative set of scene background models will switch automatically to represent the current background. This algorithm is a milestone contribution in this area, which addresses and solves various problems in the background modeling.

3.1.3 Histogram based models

In [15], Kumar et al. proposed a histogram based background modeling algorithm, to model busy scenes. Here they used a queue concept with three channels of color model Y,Cb and Cr. When the queue gets full, a calculation would be performed to regenerate the background model. This mechanism is called blind updation. The pixel values are plotted on a histogram, the values which belong to background will take the higher values in the histogram. The value of W(Gaussian mixture coefficient), is set to 5 if the background has motion due to moving branches, twinkling surfaces etc,. Each pixel will be modeled using adoptive mixture of Gaussians.

In the segmentation level, a pixel considered to be part of the foreground, when the current value of the pixel is far from the mean relative to the variance of Gaussian mixture.

3.1.4 Non-parametric model

In [4],Elgammal et al. described a basic background model and a subtraction process. The main objective of this algorithm, is to capture the information about the scene region. A continuous updation of the background information, to capture the fast changes in the background has also been proposed by this method. Here $(x_1, x_2, \ldots.., xn)$ are the recent intensity values of a pixel. Using this method, a PDF(probability Density Function) can be estimated. The non-parametric equation to estimate a pixel intensity x_t at time t, is given below:

$$Pr(x_t) = \frac{1}{n} \sum_{i=1}^{N} K(x_t - x_i) \tag{3.2}$$

Here K is the kernal estimator function, it can be calculated using a normal distribution. A pixel is considered as foreground pixel, if the probability $Pr(x_t) < th$. Here th is the global threshold, which can be adjusted to achieve the desired percentage of false positives. Unlike Gaussians, it quickly forgets the past and concentrate on the current history of pixel values.

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. Due to this factor, the input frames from distant past do not have any effect on the current background model. Compared to the non-recursive techniques, recursive techniques require less storage. However, any error in the background model can linger for a longer period of time. Some of the recursive techniques are described below:

3.2.1 Approximated median filter

To slightly improve the median based method proposed in [19]. In [2], Chung et al took the confidence values of histogram, to identify the stable intensity of each pixel over a training period. Similar method was adapted in [20], for developing the initial background model before implementing a shadow removal technique while segmentation of foreground objects in HSV color space. In [14], Kornprobst et al. assumed, that background can be defined as the most often observed part over the sequence. They have presented an approach to deal with background reconstruction and motion segmentation based on Partial Differential Equations(PDE). The results of this method were good, but choosing the parameters is difficult. Based on the same idea Hou et al. also proposed a method in [12]. In this approach, pixel intensity difference between inter-frames is calculated, and pixel intensity values are classified based on this difference. The process is performed at 4 different steps. In the initial step, classification the intensities with stable intervals and calculate the average of intensities at each interval. In the second step, classify intensities of stable intervals, with close average intensity values and take the count of the pixels in that group. Finally select the intensity value, with maximum number of pixels with homogeneous interval as background intensity value. The simulation results were good using this approach.

3.2.2 Kalman filter

A Kalman filter based mechanism was proposed in [24], which estimates the expected pixel value based on the previous observations. It explicitly considers the noise on the measurement of the system values. The components include measurement matrix, system input and Kalman gain matrix in predicting the future observation. The system estimates the values recursively and will assign the weights according to the accuracy of prediction. This model is good, in handling the illumination changes in the scene model. However, it involves higher complexity in updating.

3.2.3 Mixture of Gaussian(MoG)

This is a popular method in background modeling, it was first proposed in [5] for background modeling. They have implemented this method to explicitly classify the pixel values into three separate states corresponding to road, shadow color and colors corresponding to vehicles. Another method proposed by Stauffer et al, in [25] has drawn the attention of many researchers. In this approach rather explicitly modeling the values of all the pixels as one particular type of distribution, they simply model the values of a particular pixel as a mixture of Gaussians. Based on the persistence and the variance of each of the Gaussians of the mixture, they determine which Gaussians correspond to background colors.

Their method contains two significant parameters α - the learning constant and

T- the proportion of the data that should be accounted for by the background. In online mixture model, they consider the series of pixel values for a particular time as a "pixel process". These are scalars for gray images and vectors for color images. The following equation represents a pixel $\{x_0, y_0\}$ at time t, where I is the intensity of the image sequence.

$$\{X_1, ..., X_t\} = \{I(x_0, y_0, i) : 1 \leq i \leq t\} \tag{3.3}$$

Here $X_1..X_t$ are the sequence of images in a time frame. Different guiding factors were taken into consideration while modeling and updating the background. The following equation shows the probability of observing the current pixel value from the recent history and models it using K mixture of Gaussian.

$$P(X_t) = \sum_{i=1}^{K} w_{i,t} * \eta(X_t, \mu_{i,t}, \Sigma_{i,t}) \tag{3.4}$$

where K is the number of distributions and $\omega_{i,t}\, \mu_{i,t}, \Sigma_{i,t}$ are the estimates of weight (what portion of the data is accounted for by this Gaussian), mean value and covariance matrix of the i^{th} Gaussian in the mixture at time t. Where η is a Gaussian probability density function as shown below.

$$\eta(X_t, \mu, \Sigma) = \frac{1}{(2\Pi)^{\frac{n}{2}} \mid \Sigma \mid^{\frac{1}{2}}} e^{\frac{1}{2}(X_t-\mu_t)^T \Sigma^{-1} (X_t-\mu_t)} \tag{3.5}$$

K determined by the available memory and computational power, currently 3 to 5 are used. The covariance matrix is assumed to be of the form

$$\Sigma_{k,t} = \sigma_k^2 I \tag{3.6}$$

this assumes that the red, green, blue pixel values are independent and have the same variance. Prior to the pixel classification the weights are assigned to K distributions at time t.

A pixel in every image frame has an image measurement vector of the form $[I, I_x, I_y, I_t]$ was proposed in [23]. The advantage of having multi-dimensional Gaussian distributions allows a greater freedom to represent the distribution of the measurements occurring in the background. This concept was based on the method proposed in [9], called TAPMOG(Time Adaptive Per-Pixel MoG), where each pixel observation consists of a color and a depth measurement. The color is represented in YUV space, which helps to eliminate luminance and chromaticity, the depth measurements are denoted by D. The observation of a pixel at time t is written as $X_{i,t} = [Y_{i,t}, U_{i,t}, V_{i,t}, D_{i,t}]$.

The history of observations at a given pixel is $[X_{i,1}, ..., X_{i,t-1}]$modeled by K gaussian distributions. K is same for all pixels and it is determined to be between 3 to 5.

$$(X_{i,1}, ..., X_{i,t-1}) = \sum_{k=1}^{K} \omega_{i,t-1,k} * \eta(X_{i,t}, \mu_{i,t-1,k}, \Sigma_{i,t-1,k}) \tag{3.7}$$

The new pixel values are compared with the existing ones to find a match between the current pixel value with the existing Gaussian η_k. If a match is found, then its parameters are updated using current observation. If not, it will be replaced as a new observation. Although this method is robust with respect to foreground segmentation and background modeling, it has slow learning rate and high computational resources.

In [30], a method named 'PixelMap' was developed based on Gaussian Methodology. The motivation behind their work is, in a fast varying background few Gaussians may not be helpful, so they have extended their scope from pixel to region and frame levels. As explained before each pixel is modeled using K Gaussians categorized based on these characteristics $[Mean_{rgb}, Var_{rgb}, Max_{rgb}$ $, Min_{rgb}, Flag, Time]$

At pixel level, each incoming pixel is checked against the K Gaussian until a match is found. The moving objects have the larger variance than the background pixels. So these distributions are ordered based on decreasing order of weights. In the frame level processing each frame is subtracted from previous and next frame. The pixels which are identical in the three frames are treated as background. And a mask is obtained from this process is treated as foreground. In region level process, a shadow eliminating mechanism and a foreground connected components method is implemented to remove shadows and fill gaps in foreground regions. This method showed fairly good results, but the computation cost is high.

The improved version of Gaussian Mixture was proposed in [32]. This method adaptively updates the background, with respect to newly introduced objects and lighting conditions. The training set has been updated in reasonable time period T and at time t they maintain a data set $X_T = \{x^t, ..., x^{t-T}\}$. For each new sample data set, X_T the background will be re estimated as $\hat{p}(\vec{x}|X_T, BG)$. However, among the samples from the recent history there could be some values, that belong to the foreground objects. The estimate of those objects can be done as follows.

$$\hat{p}(\vec{x}|X_T, BG + FG) = \sum_{m=1}^{M} \hat{\pi}_m \eta(\hat{x}; \vec{\hat{\mu}}, \hat{\sigma}_m^2 I) \tag{3.8}$$

where estimates of means $\hat{\hat{\mu}}_1 ... \hat{\hat{\mu}}_M$ and estimates of the variances $\hat{\sigma}_1 ... \hat{\sigma}_M$ that describe the Gaussian components. A series of recursive update of equations were presented for this purpose. The clustering of samples enable to identify the background samples from the foreground. After identification of background clusters, they must be arranged in descending order of weights for future process. The value of Gaussian is set to 4 while experiments.

Above all are the various Gaussian methodologies that were proposed in the literature. Most of the above described methods work reasonably outdoor video surveillance systems. But the main drawback of Gaussian models are the time complexity. The per pixel processing of each image frame, consumes enormous computational resources in real time.

3.2.4 Statistical methods with combination of MoG

In [13], Kim et al, proposed a statistical model with the combination of Gaussians in their process. In this model, X is a training set for a single pixel consisting of N RGB-vectors.

$X = \{x_1, x_2, ..., x_N\}$ and $C = c_1, c_2, ..., c_N$ represents the codebook consisting of L codewords. Each pixel has different codebook size based on its sample variation. Each codeword consisting of $c_i, i = 1...L$, consists of an RGB vector $v_i = \{\bar{R}_i, \bar{G}_i, \bar{B}_i\}$ and a six tuple $aux_i = \langle \check{I}_i, \hat{I}_i, f_i, \lambda_i, p_i, q_i \rangle$, this tuple contains intensity values and temporal variables described below.

$\check{I}, \hat{I}$ - the min and max brightness, respectively

f - the frequency with which the codeword has occurred

λ - the longest period since this code word has occurred

p, q - first and access times of each codeword that has occurred.

Based on the color and brightness conditions we assign the codewords for each pixel. After constructing the code book the next step would be eliminating the foreground codewords from the constructed codebook. The new code book, after this temporal filtering, will satisfy the following condition.

$$M = \{c_m | c_m \in C \wedge \lambda_m \leq T_M\} \tag{3.9}$$

where a threshold T_M is equal to half the number of training frames, $\frac{N}{2}$. A probabilistic model using Bayesian decision theory was proposed in [16]. At the lowest level , video background segmentation is a binary classification problem. A decision to be made for each pixel in the frame at time t as *foreground* or *background.* From Bayesian perspective, a pixel being background $P(B|x)$, where x denotes the pixel observed in the frame at time t and B denotes the background class. To make this decision a gaussian should be underlying the process and the classification as follows:

$$P(x) = \sum_{k=1}^{K} P(G_k)P(x|G_k) = \sum_{k=1}^{K} W_k.g(x, \mu_k, \sigma_k) \tag{3.10}$$

where $G_k$is the k-th Gaussian and $g_k(x) \equiv g(x, \mu_k, \sigma_k)$ is the normal density function. Segmentation in this approach consists of two independent problems: estimating the distribution of all observations at the pixel level as a Gaussian mixture and evaluating how likely each Gaussian in the mixture being background. They have also proposed a generalized formula to eliminate the discontinuities of Gaussians switching from background to foreground during the process. Recently in [17], a unimodel gaussian distribution with some updation equations for foreground segmentation was proposed.

Addressing the problems with the various problems in outdoor surveillance Horprasert et al. proposed a color model that separates the brightness from chromaticity [10]. Each pixel was expressed in 3 dimensional RGB space, where $E_i = [E_R(i), E_G(i), E_B(i)]$ are expected color value of a pixel i in the reference or background image. $I_i = [I_R, I_G, I_B]$ are the intensity of the pixel in the current image. The measure of distance between the expected value and the current value using the following equation gives the brightness (α) and color distortion(CD).

$$\phi(\alpha_i) = (I_i - \alpha_i E_i)^2 \tag{3.11}$$

where α_i represents the pixel's strength of brightness with respect to expected value,

$$CD = ||I_i - \alpha_i E_i|| \tag{3.12}$$

They have considered the following attribute vector $\langle E_i, S_i, a_i, b_i \rangle$, where E_i is the expected color value,s_i is the standard deviation of color value and a_i is the variation of brightness distortion and b_i is the variation of chromaticity distortion. This method is able to deal with various outdoor conditions but the updating of the model is poorly addressed. Instead of traditional pixel based background modeling, a novel approach was presented in [31]. This method based on frequency domain analysis.

Each frame is divided into 8X8 blocks, and two features f_{DC}, f_{AC} extracted from the DCT coefficients and each block is individually modeled using single Gaussian model with mean and standard deviation (μ, σ). These two attributes are initially estimated from previous frames, possibly with some moving objects. The advantage of this kind of background modeling is, it is not sensitive to pixel noise scene changes with respective to lighting.

3.2.5 Other Statistical models

In [22], Pan et al. proposed a statistical model from first N background frames. In the modeling phase, they have considered the Mean and standard deviation of each pixel in the respective Y,Cb,Cr color space. The segmentation of foreground followed by the frame differencing. While segmentation, they have used a characteristic gain in the commercial camera and based on that factor a unimodel Gaussian was employed to segment the foreground pixels from the background. A different statistical model was presented in [8], they assumed, the pixel's intensity distribution is bimodal. The background scene is then modeled by three values, minimum pixel intensity m(x), maximum pixel intensity n(x) and the maximum difference d(x), between the consecutive frames observed during the training period. This process held in two stages, in the first stage use the median filter to distinguish the moving pixels from the stationary ones, in the second stage those stationary ones are processed to construct the initial background model. Let V be an array containing N consecutive images$V^i(x)$ is the intensity of a pixel location x in the i^{th} image of V. The initial background model is obtained as follows:

$$m(x) = min_z\{V^z(x)\}; n(x) = max_z\{V^z(x)\}; d(x) = max_z\{|V^z(x) - V^{z-1}(x)|\} \tag{3.13}$$

where $|V^z(x) - \lambda(x)| > 2 * \sigma(x)$ Hence,$V^z(x)$ is classified as moving pixel.

3.3 Conclusion

This chapter gives an brief description of various methods proposed in literature. Every method has its significance in modeling the background and segmentation. However, each method has its own drawbacks to be functional in real-time.

Chapter 4

Dynamic Background Modeling and Segmentation

This chapter gives a detailed explanation and contribution of this thesis. A histogram based background modeling algorithm is presented, which dynamically models the background with no prior knowledge of the scene. A segmentation technique is also presented, which segments the foreground objects from the background [43]. The subsequent sections present the methodology and algorithms, results can be found in the later chapters.

4.1 Dynamic Background Modeling

The key assumption in developing this method is because of increasing automation in this area. Instead of using a off line picture as a starting point, if we can able to develop a model, which can model the background by itself without user intervention would be much robust and serve the purpose in real time under different circumstances. We have also experimented widely in different color spaces, to find the suitable and reliable color space, which can handle noise and shadows at a considerable level.

HSV stands for Hue, Saturation and Value, which can also be treated as Hue, Saturation and Brightness (HSB). HSV color space created by Alvy Ray Smith in 1978. This is a nonlinear transformation of RGB color space. Possibly used in color progression. Our motivation to use this color space came from [39]. Compared to the other color spaces, it is capable to deal with noise and shadow in the image region very effectively. A pictorial representation of HSV color space is given in Fig. 4.1.

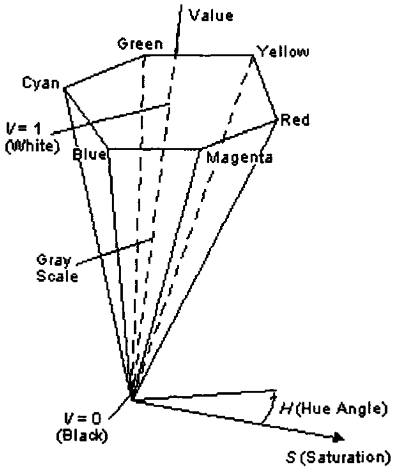

Figure 4.1: Representation of HSV color space in a 3 dimensional space [53]

In Fig 4.1, S(saturation) and V(value) parameter values have a range of [0,1] and the H(hue) parameter has a range of [0,360 degrees]. In the figure 4.1, V is the axis pointing up in the picture. With V=*1*, at the top represents relatively bright colors. Due to this unique separation of brightness property, we used HSV space rather the traditional RGB space.

Many other works in HSV color space can be found in [37], [38],[33].

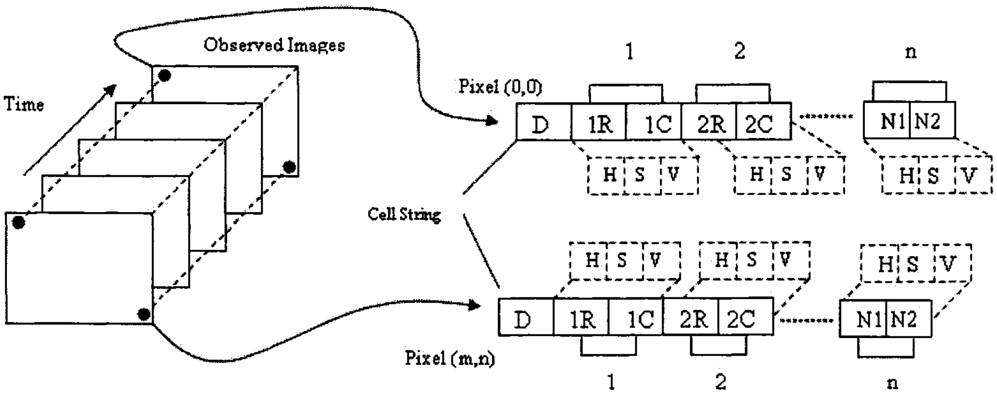

Figure 4.2: History map: A sample overview of how pixels are being captured over time

As shown in Fig 4.2, the *history map* represents intensity characteristics of a pixel over a time period. The *observed images* are the image frames obtained from the video over a period of time. The *cell string* represents entire intensity variation at each pixel location over all the frames. Each pixel has three components *H*,*S* and *V*. The information can be used to estimate the variation of pixel intensity over a time period. The conclusions drawn from these, can be used for the identification of stable intensity pattern over time. Several methods had been proposed underlying this concept. The simple method to start with is the median calculation of this interval. Some other methods namely estimation of Gaussian coefficient or Gaussian mixture.

In the present context, we have used a histogram based method. This histogram method identifies the intervals of stable intensity of each pixel. Later, these pixel intensities for each pixel are back substituted to form an approximate background model.

4.1.1 Histogram based Background Model

In general, a histogram displays continuous data in ordered columns. It an effective way to represent the continuous measures, such as time, intensity etc. Histogram is the best way, to identify the stable intensity of each pixel in certain time interval.

The histogram in this case is formed of HSV components, for each individual component H,S and V, we have identified a stable intensity occurred most of the observed period of time and took into account. The values are separated among four bins, we have chosen four bins because of the stability in results. Values in each bin represents the color information(*intensity*)at different time intervals. The same procedure has been applied for all the pixels. The procedure for the histogram based method given below:

1. Convert each frame in HSV color space.
2. Calculate the histogram of H, S and V component for a particular pixel in all the fames.
3. Find the histogram bin with highest value and assign the median of this bin to the H, S and V component of the pixel in the background model.
4. Perform step 1,2 and 3 for all the pixels in the background model until the pixel values are stabilized.

Table 4.1: Histogram based Background Modeling

A sample graph is formed of 'V' component of the HSV color space in Fig 5.3. The horizontal axes represent the intensity value and vertical axes represents the count of each bin.

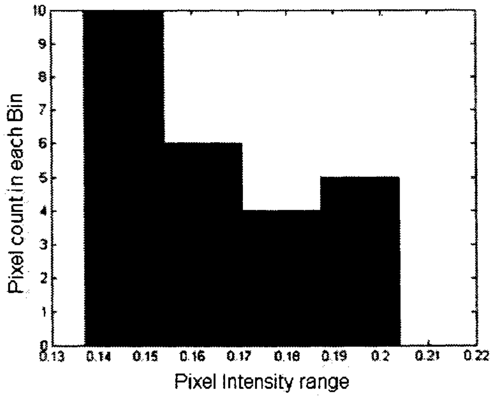

Figure 4.3: Sample histogram formed of 25 image frames on V component

Once the background model has been identified, the next step is to segment the foreground moving objects from the background. In many previous methods, an approximate threshold value is being used to differentiate the foreground pixels from the background. However, setting a threshold value is not so robust and may effect when the scene changes dynamically. Because of this reason, we have used K-means clustering technique, which automatically calculates the threshold value. The algorithm and explanation can be found in the subsequent chapters.

4.2 Segmentation by Clustering

Clustering is a classification technique. In measurement space it is used as an indicator of similarity of image regions. It may also be used for segmentation purposes. Similarity between image regions or pixels can be grouped, by using clustering (small separation distances) in the feature space. These clustering methods come under the earliest data segmentation techniques.

In Fig 5.4, a sample visible partition of two different groups using K-means clustering can be seen. K-Means clustering finds, a grouping of the pixels which are similar in color, position or a combination of both etc. The similarity measure between the feature vectors is calculated by the Euclidian distance. Here the value of 'K' represents number of clusters to be formed among the pixels of the image. While

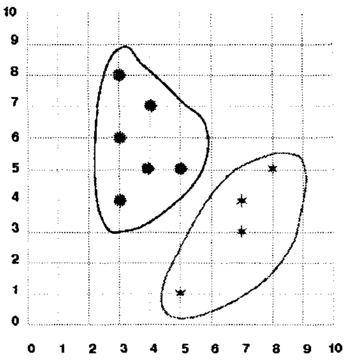

Figure 4.4: Sample clustering mechanism using K-means [54]

implementing this technique, we have used the connected components approach to merge the group of individuals, whose size is less than some threshold to the adjacent component.

In [5], a clustering algorithm was proposed based on the intensity values. In this technique, the pixel intensity is observed over a period of time then it is classified either a foreground or a background. The K-clusters model consists of K weights and K average pixel values or centroids. In this approach, the user has to specify the threshold limit for pixel before assigning to a specific cluster. If a pixel do not fall in an existing cluster, then a new cluster is created for that particular pixel. This algorithm uses Manhattan distance to identify whether a pixel is above or below the specified threshold. In particular, this method can not identify the difference between the moving objects and shadows. The manual specification of the threshold might not be a robust idea as this paper brought forth.

4.2.1 K-Means Clustering technique

We initially subtracted an image frame from the background model, which gives us a difference matrix. From the difference matrix, we took maximum and minimum intensity values and named them as Meanl and Mean2. Later, we have applied the K-means clustering algorithm on the difference matrix to classify pixels into two groups, to segment foreground pixels from the background pixels.

In this case, K is set to 2, which represents two clusters one for the foreground and the other for the background. The algorithm is given below:

1. Convert video frame into HSV color space.

2. Subtract the H, S and V component of background model[obtained using Table 5.1] from the H,S and V components of the video frame and store the absolute values into a difference image.

3. Find minimum and maxim values in the difference matrix. The minimum value corresponds to the seed of the background cluster, denoted by M1 and maximum value corresponds to the seed of foreground cluster, denoted by M2.

4. If $D_{i1} > D_{i2}$, then assign i to FG cluster, otherwise assign i to background cluster.

5. Calculate mean of background cluster, M1 and mean of foreground cluster, M_2.

6. Repeat step 4 and 5 until $M_1$and M_2 does not change significantly
(The difference between the two mean should be approximately zero).

7. Report pixels in foreground cluster as foreground region and vice versa.

Table 4.2: Segmentation by k-means clustering

4.2.2 Problems in Segmentation

Segmentation suffers from different problems. While segmenting the regions of foreground, regions similar to the background will not appear after the segmentation process. This situation is difficult to handle in traditional RGB space. However, in HSV color space it can be handled by using chromaticity information represented by Hue and Saturation.

In our experiments, we were able to solve this problem by observing the color region and setting a measure so that these image regions are considered as foreground in the final result. We considered the pixels, whose saturation component value is greater than some minimum value and hue value is not zero. This is because all pixels with saturation zero will appear same in color [40]. When the saturation component of the pixel increases, the color value changes, which in turn changes the hue value (from zero to some particular value). This characteristic allows us to search for a value, by which we could extract those pixels from the segmented foreground regions. Needs to be cautious of not including one the pixels from shadows around the foreground regions. This is also depends on the value we decide.

We have also tried to increase the number of clusters while segmentation, but the results with two clusters were fairly good enough for decent segmentation.

4.3 Conclusion

We have proposed two main techniques, which contribute to the first building block of the video surveillance system. We have presented a histogram based method, which models the background dynamically in busy scenes. This kind of model suitable to model scenes in different places, which includes shopping malls, indoor building environments etc. We have some limitations in using this method, it creates blended regions if the scene is too busy. We have also presented a segmentation technique, which uses K-means clustering. K-means clustering is a popular method in image processing applications, we limit the number of clusters(k) to 2. We found two major problems in segmentation phase, which are due to shadows and foreground aperture. The detailed discussion can be found in the above sections. The implementation and results are presented in the subsequent chapters.

Chapter 5

Implementation and Results

We have implemented the proposed methods in MATLAB version 7 with Image Processing Toolbox. This machine runs on a Pentium 4 processor, having 2.79GHz bus speed. We have acquired the images(dimensions:320X240) from a movie shot by an off-the-shelf digital camera. Initially, we considered the burst mode to capture pictures. Later on, we have tested these methods successfully on the frames extracted from the video.

Throughout our experiments, we have assumed constant lighting conditions all across the frames. We have also assumed there is no movement of the camera while shooting the videos. The processing speed of our algorithm is 2-4 frames/sec(MATLAB), which can be further improved after optimization. We have widely experimented our methods in different indoor environments.

5.1 Background Modeling using Histogram

As mentioned in the previous chapter, we have used the histogram of four bins in our method. The pixel intensity values are scattered among the four bins. The histogram of maximum number of pixels will be considered for finding the background pixel intensity.

In Fig.5.1, a sample set of frames were shown. These frames extracted from the video taken in a corridor. This is a typical environment in video surveillance. The result of this set of images can be seen in Fig.5.2. This result is almost perfect to the ideal background image.

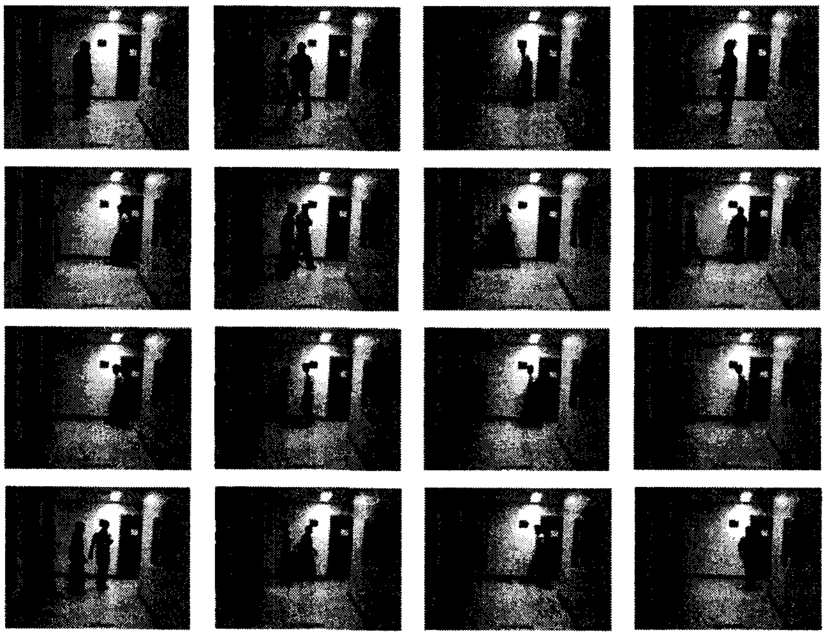

Figure 5.1: A series of images extracted from the video took inside the building

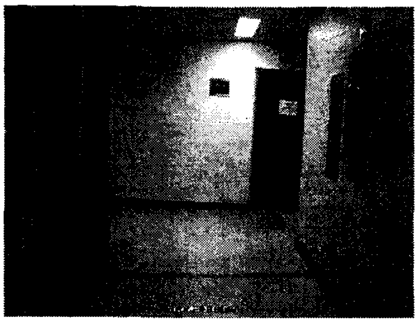

Figure 5.2: Identified background of the series shown in Fig 5.1

In Fig 5.3, the image set is extracted from a video taken in a busy shopping mall. Finding a background in these situations is little difficult. Which due to the fact that, the scene is busy at all times. The resulting background of this set can be seen in Fig.5.4. This result is almost resembles the original background, but it has some blurred areas. These blurred areas are caused by blending of the image pixels. The detailed discussion can be found at the end of this section.

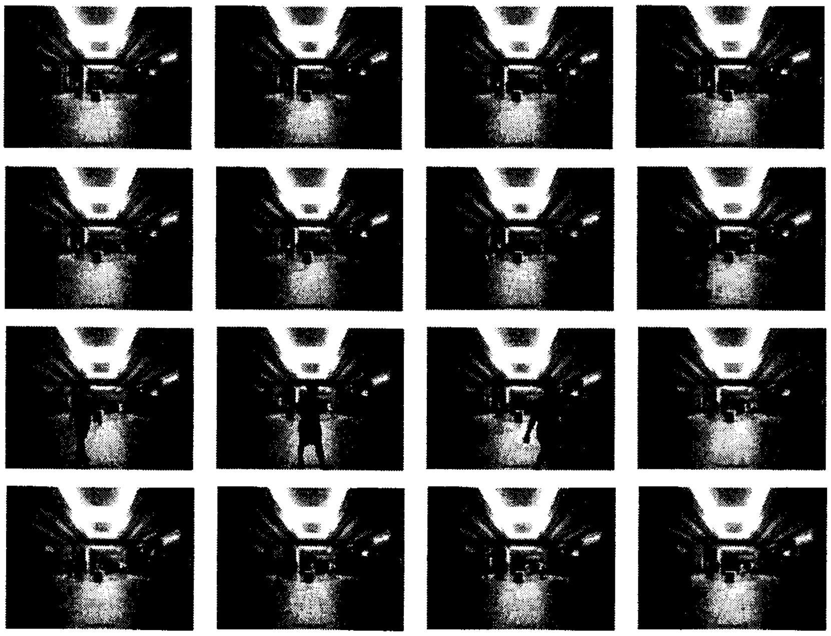

Figure 5.3: Image series of the video took inside the shopping mall

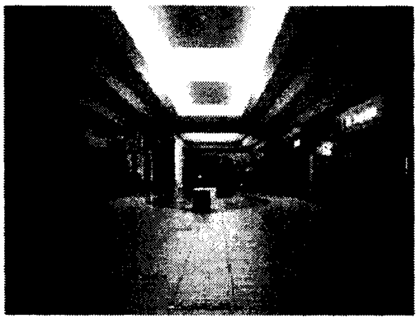

Figure 5.4: Identified background of the series shown in Fig 5.3

In Fig 5.5, another set of images taken in the mall environment. In this case camera mounted at a corner. The corners were crowded almost every time. The resulting background of this series can be seen Fig 5.6. We can see some regions seemed to have blending of pixels.

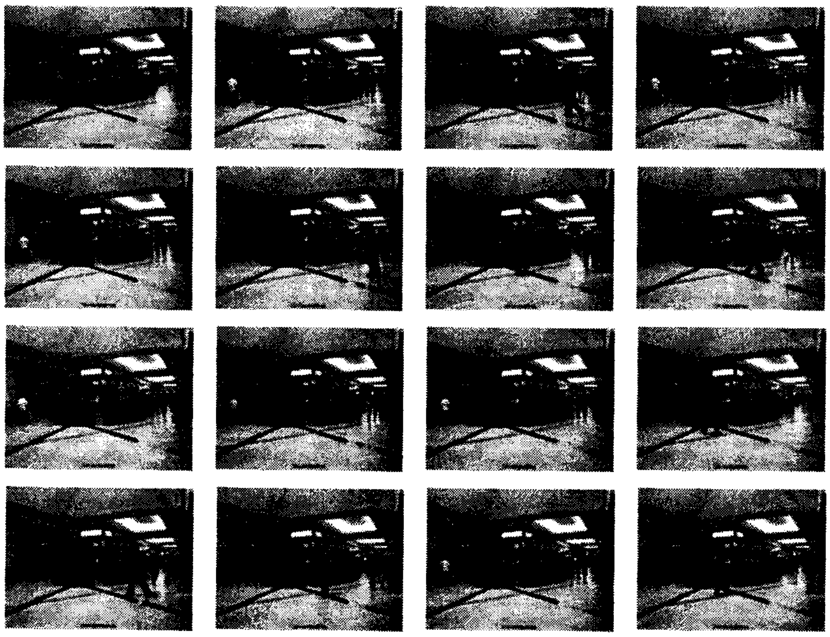

Figure 5.5: Another image series of the video inside the shopping mall

Figure 5.6: Background of the series 5.5

5.1.1 Discussion

Blending is a common problem in modeling the background scene. Blending occurs mainly because of no stable pixel intensity at any specific region. In the above results, blending is observed in couple of places due to the busyness of the scene. This is hard to avoid, but can be solved by better statistical analysis.

The stabilization of the background can be reported when one bin out of four bins reaches 50% of the total probability. This measure can be adopted under any situation to declare a stable background has been achieved.

5.2 Segmentation by K-means Clustering

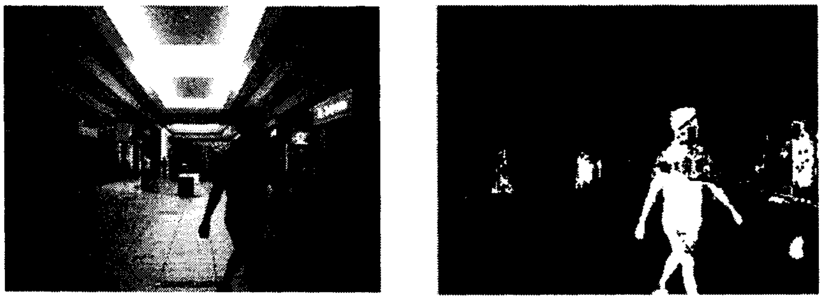

Figure 5.7: Result of segmentation in a shopping mall

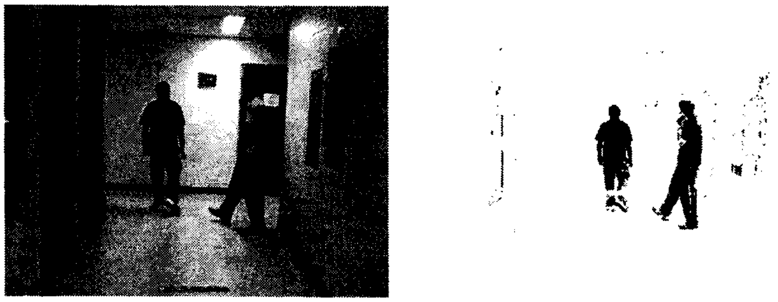

Figure 5.8: Result of segmentation in a corridor

In Fig.5.7 and Fig.5.8, the result of the K-means clustering technique in different situations was given. We picked some noise while segmentation, most of the noise was considerably eliminated by our method and the color space we used. However, still some areas are partially detected falsely due to heavy noise in the video. Initially, we used the value (V) of HSV space, later we used the chromaticity information (H and S) for increasing the efficiency of segmentation result. In these experiments, we faced a problem called foreground aperture. This problem occurs, if the region of foreground is more similar to the background. In that case, areas having similar features will be ignored in the final result. The details of this problem can be found in discussion section.

5.2.1 Discussion

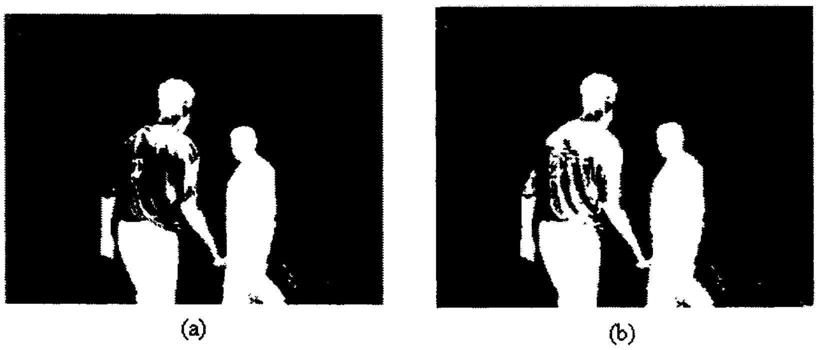

Figure 5.9: Problems in Segmentation: (a)Partially segmented part of shirt,(b)almost detected region after considering chromaticity information

Figure 5.10: Experimented Image of the above results

Foreground aperture is a traditional problem in segmenting the foreground. We have used, a simple solution which can detect the missing foreground regions. After segmenting the foreground region using V, we used H and S components to detect the missing regions. The prime cause of missing regions in the foreground is due to the following reasons:

1. A close match of foreground region to the background, which is called foreground aperture.

2. Shadows in the image region, basically there are two kinds shadows identified as problematic. They are self shadows and cast shadows. Cast shadows should be eliminated while segmentation but self shadows should be considered as they are part of the foreground regions. But in the present experiment 5.9 we failed to segment the self shadows due to their close texture with the background color. Self shadow is the shadow cast by the object on the foreground region.

These false positives can be detected partially using the complete color information provided by the color space (limited to these experiments), but can also be detected using size and distance also.

In figure 5.9 (a) the shirt region of the first person is partially missing (b) the missing region is almost detected using the chromaticity information. But some parts still have missing due to above mentioned problem of self shadows.

In Fig.5.11, a comparison of present methods to the literature methods. The data set is obtained from the authors of [28]. The proposed method, performed considerably better than some of the literature methods.

5.3 Conclusion

This chapter shows the results of successful implementation of our methods. The initial section shows the results of statistical modeling of background using histogram. A discussion is being presented on the blending problem in background modeling. In the subsequent sections, we presented our results of our segmentation strategy, we have also addressed some problems in segmentation in those sections. At the end of this chapter, we have presented a comparison of our method with the literature methods.

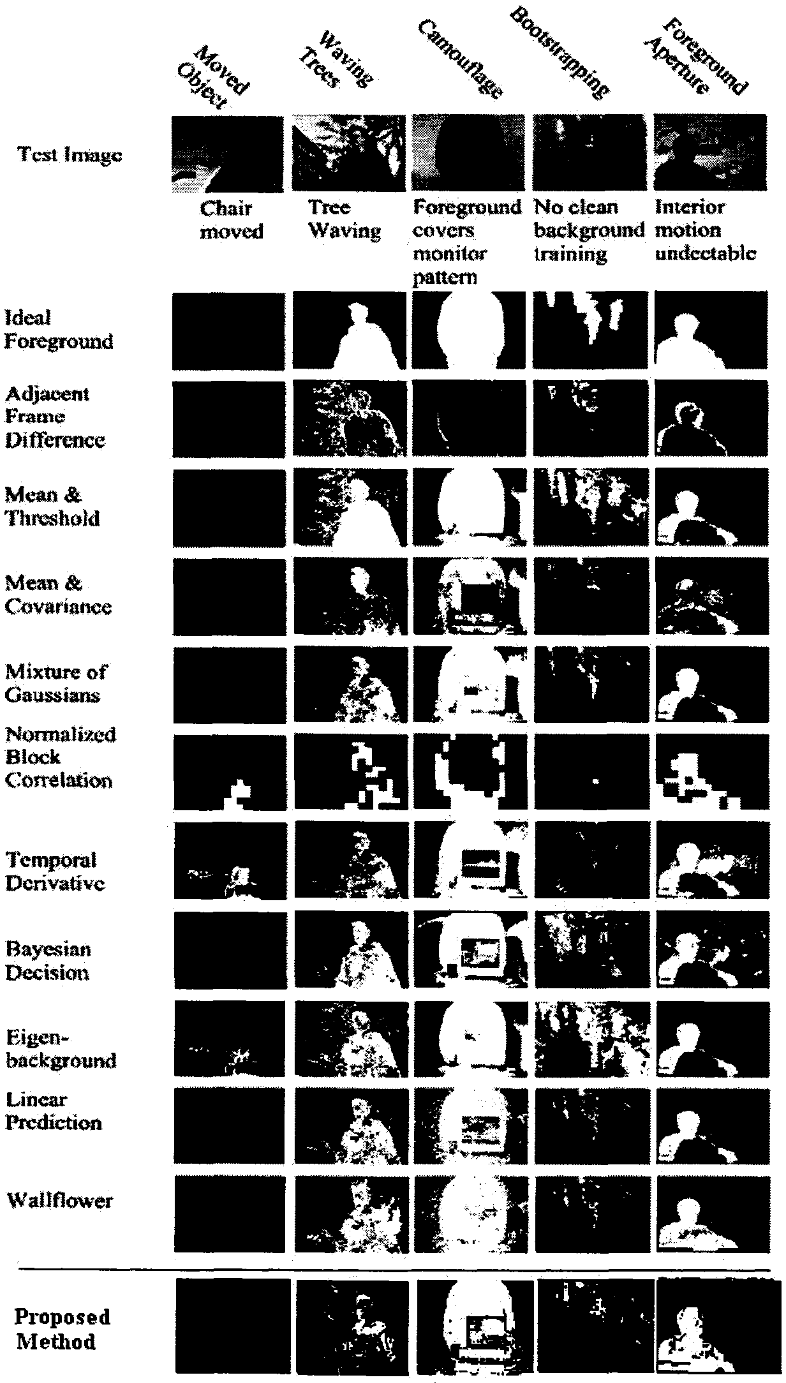

Figure 5.11: The present method is compared to different other methods as mentioned in Walflower[28] algorithm

Chapter 6

Conclusions

Video surveillance has become a popular research area in the field of Computer Vision. The reason behind this advancement is due to inexpensive computational and hardware resources, and growing concern for security and surveillance in almost every place we visit. Automation of this system involves many issues to be taken care of, which starts from building a background model with no foreground objects, segmenting the moving foreground objects and tracking those objects to identify the suspicious behavior.

This thesis motivated from a general perception, of how to build a background model when the objects present in the scene. Keeping this concern in mind, we have named it dynamic background modeling. Dynamic background modeling can be defined as, modeling a scene dynamically with no moving objects under any situation with no prior knowledge. In order to accomplish this task, one has to track the history of the scene background and analyze this history in some intelligent fashion, to get the information about the scene background.

We have used a histogram based method to analyze the background history, and used this information to build the approximate background model. We have used four bins, to analyze each pixel's history over a time period to find an intensity value, which represents the original background pixel with no influence of foreground objects. However, finding this value can be difficult, due to continuous motion in the scene region. A better statistical analysis might improve the results in this case. Systems in real time need fastness, we claim this method works fairly fast enough with considerable robustness in different situations.

The next step after finding a background is, segmenting the moving objects in the scene. Many segmentation mechanisms work on approximate threshold value, this decided value will be used to segment the pixels belong to the foreground. However, this kind of analysis requires human observation. We dealt this situation, by using a clustering mechanism, which calculates the threshold value to separate the foreground objects in an effective manor. We have used two clusters for this purpose, we relied primarily on the color information of images. Shadows and noise are two evident problems while segmentation, but we make use of brightness and chromaticity information to deal these situations. Throughout the experiments, we have converted the traditional RGB images into HSV color images. As mentioned in the previous chapters, HSV color space works relatively better to deal with noise and shadows.

In conclusion, this work contributes mainly in segmentation area of the video surveillance system. Segmentation plays a crucial role in later stages of the system, but proper segmentation depends on various factors. An intelligent system should able to support itself, to decide the value on which it can segment the foreground objects. This method can suffice the cause with in reasonable bounds.

We have modeled the background using a histogram based method, but it is statistically fragile, when considered the situations like continuous object flow in the background. Situations like these need further concentration, by which possibilities of blending can be avoided. Many previous methods suffer from this problem, present method is not an exception in some cases.

Segmentation is a broad and rather a difficult area to deal with. It needs concentration in handling noise and shadows effectively. Self shadows are difficult to deal with which we failed to accomplish in our work. Relying on mere color information might be problematic in some situations. The use of geometrical information to recognize the missing parts of the foreground objects is necessary. However, clustering mechanism might fail, if the foreground object's color matches close to the background. Use of other clustering techniques like agglomerative, hierarchical etc., are still unexplored for the segmentation purpose.

Bibliography

[1] Sen-Ching S. Cheung and Chandrika Kamath. Robust background subtraction with foreground validation for urban traffic video, *"EURASIP Journal on Applied Signal Processing "*, 2005:2330–2340, January 2004.

[2] Y.C. Chung, J.M. Wang, and S.W. Chen. A vision-based traffic light detection system at intersections, *"Journal of Taiwan Normal University: Mathematics, Science and Technology"*, 1:1:67–86, 2002.

[3] K. Dwason-Howe. Active surveillance using dynamic background subtraction, *"Technical Report TCD-CS-96-06, Trnity College"*, 1996.

[4] A. Elgammal, D. Harwood, and L. S. Davis. Non-parametric model for background subtraction, *"In European Conference Computer Vision"*, 2:751–767, 2000.

[5] Nir Friedman and Stuart Russell. Image segmentation in video sequences: a probabilistic approach, *"Uncertanity in Artificial Intelligence"*, pages 175–181, 1997.

[6] B. Gloyer. Video-based freeway monitoring system using recursive vehicle tracking, *"In Proceedings of IS and T-SPIE Symposium on Electronic Imaging: Image and Video Processing"*, 1995.

[7] D Gutchess, M Trajkovics, E Cohen-Solal, D Lyons, and A.K Jain. A background model initialization algorithm for video surveillance, *"Proceedings of Interrnational Conference on Computer Vision"*, 1:733–740, July 2001.

[8] Ismail Haritaoglu, David Harwood, and Larry S. Davis. Fast background scene modeling and maintenance for outdoor surveillance, *"International Conference on Pattern Recognition"*, 4:179–183, Sept 2000.

[9] M. Harville. A framework for high-level feedback to adaptive, per-pixel, mixture-of-gaussian background models, *"European Conference on Computer Vision"*, 3:543–560, May 2002.

[10] T. Horprasert, D. Harwood, and L. S. Davis. A robust background subtraction and shadow detection, *"Proceedings of Asian Conference on Computer Vision"*, 2000.

[11] Thanarat Horprasert, David Harwood, and Larry S. Davis. A statistical approach for real-time robust background subtraction and shadow detection, *"International Conference on Computer Vision"*, 1999.

[12] Zhiquiang Hou and Chongzhao Han. A background reconstruction algorithm based on pixel. intensity classification in remote video surveillance, *"International Conference on Information Fusion"*, pages 754–759, July 2004.

[13] Kyungnam Kim, Chalidabhongse T.H., Harwood D, and Davis L. Background modeling and subtraction by codebook construction, *"International Conference on Image Processing"*, 5:3061–3064, Oct 2004.

[14] P Kornprobst, R Deriche, and G Aubert. Image sequence analysis via partial difference equations, *"Journal of Mathematics Imaging and Vision"*, 11:1:5–26, 1999.

[15] Pankaj Kumar, Surendra Ranganath, and Weimin Huang. Queue based fast background modelling and fast hysteresis thresholding for better foreground segmentation, *"IEEE ICICS - PCM"*, pages 743–747, Dec 2003.

[16] Dar-Shyang Lee, Hull J.J, and Erol B. A bayesian framework for gaussian mixture background modeling, *"Proceedings of International Conference on Image Processing"*, 3:III–973–6, Sept 2003.

[17] Jordi Llus, Xavier Miralles, and Oscar Bastidas. Reliable real-time foreground detection for video surveillance applications, *"Proceedings of ACM international workshop on Video surveillance and sensor networks "*, 3:59–62, Nov 2005.

[18] W. Long and Y.H. Yang. Stationary background generation: An alternative to the difference of two images, *"IEEE Transactions on Pattern Recognition"*, 23:12:1351–1359, Nov 1990.

[19] M Massey and W Bender. Salient stills: Process and practice, *"IBM Systems Journal"*, 35:4:557–573, 1996.

[20] I. Mikic, P.C. Cosman, G.T. Kogut, and M.M. Trivedi. Moving shadow and object detection in traffic scenes, *"International Conference on Pattern Recognition"*, 1:321–324, Sept 2000.

[21] Naoya Ohta. A statistical approach to background subtraction. for surveillance systems, *"International Conference on Computer Vision"*, pages 481–486, 2001.

[22] Jinhui Pan, Chia-Wen Lin, Chuang Gu, and Ming-Ting Sun. A robust video object segmentation scheme with prestored. background information, *"IEEE International Symposium on Circuits and Systems"*, 2002.

[23] Robert Pless, John Larson, Scot Siebers, and Ben Westover. Evaluation of local models of dynamic backgrounds, *"International Conference on Computer Vision and Pattern recognition"*, 2:II–73–8, June 2003.

[24] C. Ridder, O. Munkelt, and H. Kirchner. Adaptive background estimation and foreground detection using kalman filtering, *"In International Conference on Recent Advances in Mechatronics"*, pages 193–199, 1995.

[25] Chris Stauffer and W.E.L.Grimson. Adaptive background mixture models for real-time tracking, *"International Conference on Computer Vision and Pattern Recognition"*, 2:–252, 1999.

[26] Luigi Di Stefano, Giovanni Neri, and Enrico Viarani. Analysis of Pixel-Level Algorithms for Video Surveillance Applications , *"International Conference on Image Analysis and Processing"*, pages 541–546, 2001.

[27] Jen-Chao Tai and Kai-Tai Song. Background segmentation and its application to traffic monitoring using modified histogram, *"International Conference on Networking, Sensing and Control"*, 1:13–18, 2004.

[28] Kentaro Toyama, John Krumm, Barry Brumitt, and Brian Meyers. wallflower: Principles and practice of background maintenance, *"International Conference on Computer Vision"*, 1:255–261, Sept 1999.

[29] Wikipedia. computer vision: An introduction, *"www.wikipedia.org"*, 2006.

[30] Qi Zhang and Reinhard Klette. Robust Background Subtraction and Maintenance, *"International Conference on Pattern recognition"*, 2:90–93, Aug 2004.

[31] Juhua Zhu, Stuart Schwartz, and Bede Liu. A transform domain approach to real-time foreground segmentation in video sequences, *"International Conference on Acoustics, Speech, and Signal Processing"*, March 2005.

[32] Zoran Zivkovic. Improved adaptive gaussian mixture model for background subtraction, *"International Conference on Pattern Recognition"*, 2:28–31, 2004.

[33] Shamik Sural, Qian Gang, Sakti Pramanik. Segmentation and histogram generation using the HSV color space for image retrieval, *"International Conference on Image Processing"*,II 589–II 592 2002.

[34] D.M. Gavrila. The Visual Ananlysis of Human Movement: A Survey, *"Computer Vision and Image Understanding"*,73:82–98,1999.

[35] M. A. Ali. Feature based Tracking of Multiple People For Iintelligent Video Surveillance, *"Master's Theis, School of Computer Science, University of Windsor"*,July,2006.

[36] Wikipedia. Computer Vision: An Introduction, *"www.wikipedia.org"*,2006.

[37] Na Li,Jiajun Bu,Chun Chen. Real-time video object segmentation using HSV space, *"International Conference on Image Processing"*,II-85–88, 2002.

[38] Baisheng Chen,Yunqi Lei. Indoor and Outdoor People Detection and Shadow Suppression by Exploiting HSV Color Information, *"The Fourth International Conference on Computer and Information Technology"*,137–142, 2004.

[39] Alexandre F, Gerald M. Adaptive color background modeling for real-time segmentation of video streams, *"Proceedings of International on Imaging Science, System and Technology"*,227–232, 1999.

[40] Darrin Cardani. Adventures in HSV Space, *(http://www.buena.com/articles/hsvspace.pdf)*.

[41] Ying-Li Tian and Arun Hampapur. Robust Salient Motion Detection with Complex Background for Real-time Video Surveillance, *"IEEE Workshop on Applications of Computer Vision"*,30–35, 2005.

[42] L. Di Stefano, G. Neri and e. Viarani. Analysis of pixel-level algorithms for video surveillance applications, *"11th International Conference on Image Analysis and Processing"*,541–546, 2001.

[43] Simeon Indupalli, M.A. Ali, Bubaker Boufama. A Novel Clustering based Method for Adaptive Background, Segmentation *"3rd canadian Conference on Computer and Robot Vision-Workshop on Video Surveillance and Security"*,37, 2006.

[44] J Dever, N da Vitoria Lobo, M Shah. Automatic Visual Recognition of Armed Robbery, *"Proceedings 16th International Conference on Pattern Recognition"*,1:451–455, 2002.

[45] Alexandre F, M Grrald. Adaptive color background modeling for real-time segmentation of video streams, *"Proceedings of lntemational on Imaging Science, System and Technology"*,227–232, 1999.

[46] Shao-Yi Chien, Shyh-Yih Ma, and Liang-Gee Chen. Efficient Moving Object Segmentation Algorithm Using Background Registration Technique, *"IEEE Transactions on circuits and systems for video technology"*,12:577–586, 2002.

[47] Andrea Cavallaro, Olivier Steiger and Touradj Ebrahimi. Tracking video objects in cluttered background, *"IEEE Transactions on Circuits and Systems for Video Technology"*,4:575–584, 2004.

[48] Xiang Gao, T.E. Boult,Frans Coetzeey,V Ramesh. Error Analysis of Background Adaption, *"IEEE Computer Vision and pattern Recognition"*,1:503–510, 2000.

[49] J. J. Yoon, C.Koch and T. J. Ellis. Shadow Flash: an approach for shadow removal in an active illumination environment, *"British Machine Vision Conference"*,636–645, 2002.

[50] A. Prati and I. Mikic and R. Cucchiara and M. Trivedi. Analysis and Detection of Shadows in Video Streams: A Comparative Evaluation, *"IEEE CVPR Workshop on Empirical Evaluation Methods in Computer Vision"*,2001.

[51] Dana H. Ballard and Christopher M. Brown. Computer Vision, *"Prentice Hall Publication, ISBN:0-13-165316-4"*,1982.

[52] Shimon Edleman. Representation and Recognition in Vision, *"The MIT Press"*,1999.

[53] MicroSoft Development Network. HSV Color Spaces, *"www.windowssdk.msdn.microsoft.com"*,2006.

[54] WWW.togaware.com. K Means Option, *"www.togaware.com"*,2006.

Vita Auctoris

Simeon Indupalli was born in 1980 in Jaggaiahpet, Andhra Pradesh, India. He earned B.Sc(Statistics, Comp.Science) and Master of Engineering(Computer Applications) from Nagarjuna University, in 2000 and 2003 respectively. He joined the University of Windsor in September 2004 and currently a candidate for the Master's degree in Computer Science at the University of Windsor and is to graduate in Fall 2006.