

2009-01-01

Real-time Image-based Motion Detection Using Color and Structure

Manali Chakraborty

University of Texas at El Paso, mchakraborty@miners.utep.edu

Follow this and additional works at: https://digitalcommons.utep.edu/open_etd



Part of the [Computer Sciences Commons](#)

Recommended Citation

Chakraborty, Manali, "Real-time Image-based Motion Detection Using Color and Structure" (2009). *Open Access Theses & Dissertations*. 2658.

https://digitalcommons.utep.edu/open_etd/2658

This is brought to you for free and open access by DigitalCommons@UTEP. It has been accepted for inclusion in Open Access Theses & Dissertations by an authorized administrator of DigitalCommons@UTEP. For more information, please contact lweber@utep.edu.

REAL-TIME IMAGE-BASED MOTION DETECTION USING COLOR AND
STRUCTURE

MANALI CHAKRABORTY

Department of Computer Science

APPROVED:

Olac Fuentes, Chair, Ph.D.

Eric Freudenthal, Ph.D.

Sergio Cabrera, Ph.D.

Patricia D. Witherspoon, Ph.D.
Dean of the Graduate School

©Copyright

by

Manali Chakraborty

2009

to my

FAMILY

with love

REAL-TIME IMAGE-BASED MOTION DETECTION USING COLOR AND
STRUCTURE

by

MANALI CHAKRABORTY, B.TECH.

THESIS

Presented to the Faculty of the Graduate School of

The University of Texas at El Paso

in Partial Fulfillment

of the Requirements

for the Degree of

MASTER OF SCIENCE

Department of Computer Science

THE UNIVERSITY OF TEXAS AT EL PASO

December 2009

Acknowledgements

In the first place I would like to express my heart felt gratitude to my advisor, Dr. Olac Fuentes of the Computer Science Department at the University of Texas at El Paso, for his supervision, advice and guidance. His enduring patience and never ending belief in my capabilities inspired my growth both as a student and as a researcher. I could not have hoped for a better mentor and advisor. I would also like to extend my thanks to my committee members Dr. Eric Freudenthal and Dr. Sergio Cabrera whose suggestions, comments and insights were extremely helpful for the completion of this thesis.

Many thanks go in particular to my fellow lab-mates Steve Gutstein, Geovany Ramirez and Jun Zheng for using their precious time to read this thesis and providing critical feedback to make it better. Also I deeply acknowledge the contributions of my friends Joel, Nicolle, David and Brenda at the Vision and Learning Lab without whom this thesis would not have been possible.

In addition to above I would like to thank the professors of the Computer Science department at the University of Texas at El Paso whose dedication and hard work have made me the computer scientist I am today. I would especially like to take this opportunity to thank Dr. Vladik Kreinovich, Dr. Martine Ceberio and Dr. Luc Longpre for their help and patience.

Beyond the realm of Computer Science I would like to thank my old friends and roommate Shubhra for putting up with my weirdness specially in the last few days, Bivas for his knack for figuring out things in tricky situations, Somdevda for his numerous rides to the local grocery store and Ruku for his unfailing obsession with fitness programs. It would have been a different world without them.

And finally I must thank my family for their continuous support and belief in me even when I doubted in myself and Amritam for always being there for me with love and lots of bright ideas.

Abstract

Motion detection is an important problem in computer vision and has multiple applications in the real world, including surveillance-related activities, gaming, and human-robot interactions. These systems need to be robust enough to handle fluctuations in light intensity and other external factors like noise and compression artifacts. In this thesis a method is proposed for detecting the regions of motion from a video sequence in real time. The main idea of this work is to detect motion based on both structure and color. Structure-based detection is carried out using information from the Census Transform computed on gradient images obtained with Sobel operators. The Census Transform characterizes local intensity patterns in an image. Color-based detection is done using color histograms, which allow efficient characterization without prior assumptions about color distribution in the scene. The probabilities obtained from the gradient-based Census Transform and Color Histograms are combined in a robust way to detect the active motion zones. Experimental results demonstrate the effectiveness of this approach.

This thesis also presents an application for motion detection in foveal visual systems. Foveation reduces power and bandwidth requirements in a system. But the main challenge is to identify the region of interest that must be transmitted in high-resolution format while maintaining the rest in low resolution. In this work this issue is addressed, by proposing motion detection as a cue to select the region of interest in a given scene. Preliminary results are also presented to show the successful application of this approach.

Table of Contents

	Page
Acknowledgements	v
Abstract	vi
Table of Contents	vii
List of Figures	ix
Chapter	
1 Introduction	1
1.1 Motivation	1
1.2 Objective	2
1.3 Research Scope	3
1.4 Approach	3
1.5 Implementation	4
1.6 Contribution	5
1.7 Outline	5
2 Previous Work on Motion Detection	7
2.1 Temporal Differencing	7
2.2 Background Modeling	8
2.3 Optical Flow	10
2.4 Entropy Based Models	10
2.5 Direct detection	11
3 Solving Motion Detection	13
3.1 Introduction	14
3.1.1 The Census Transform	14
3.1.2 The Modified Census Transform	14
3.1.3 Temporal Color Histogram	15

3.2	Proposed Algorithm for Detection	16
3.2.1	Training	17
3.2.2	Testing	18
3.2.3	Combining Information from Color and Structure	20
4	Results and Observations	22
4.1	Experiments Conducted	22
4.2	Detected Results	22
4.3	Analysis of Experiments	24
4.3.1	Sensitivity to Low Resolution	25
4.4	Observations	28
5	Foveation	30
5.1	Introduction	30
5.2	Some Related Work on Foveal Techniques	31
5.3	Foveal Detection	32
5.4	Experimental Results	34
6	Concluding Remarks	36
6.1	Significance of the Result	36
6.2	Future Work	36
	References	38
	Curriculum Vitae	44

List of Figures

1.1	Flow of the Algorithm	4
2.1	Temporal Differencing e.g.1	8
2.2	Temporal Differencing e.g.2	8
2.3	Detection using Optical Flow Algorithm	11
3.1	Working Principle	13
3.2	Census Transform	15
3.3	Color Cube Model	16
3.4	Flow Chart	17
3.5	Testing Algorithm	20
4.1	Frames 70-86 from OutdoorScene1	23
4.2	Detected Frames 70-86 from OutdoorScene1	24
4.3	Frames 50-58 from OutdoorScene1	25
4.4	Frames 315-323 from OutdoorScene2	26
4.5	ROC curves for OutdoorScene1	27
4.6	ROC curves for OutdoorScene2	28
4.7	Comparison of ROC curves	29
5.1	Foveation in OutdoorScene1	33
5.2	Foveation in OutdoorScene3	34

Chapter 1

Introduction

Motion detection is one of the most important problems in computer vision. Motion detectors are commonly used as initial stages in several surveillance-related applications such as person detection, person identification, and activity recognition. Changes in illumination, noise, and compression artifacts make this a challenging problem. In this thesis a solution to this problem is proposed by developing a system which can effectively handle similar situations in real time. Of the numerous tools that exist for detecting motion hardware devices include embedded sensors to identify the movements of objects. Cameras with different views for recording and transmitting images can also capture and detect moving objects. However pre-processing time required to perform detection in all these systems is significantly more compared to the proposed solution in this thesis where extremely low resolution image data is used as input.

Earlier detection techniques implemented by vision scientists have taken into consideration different aspects of the system, but have failed to prove one point. This involves reducing the amount of input data to be processed in order to carry out detection at an effective rate. The low resolution approach therefore opens up new vistas in vision science and can also be extended to other real-time applications.

1.1 Motivation

The motivation of this thesis is to build a detector system which not only uses low resolution image data to do the processing but is also reliable enough to handle variations in light intensity. Specifically outdoor installations of motion capturing devices often face such

challenges where fluctuations in weather conditions has a direct impact on the performance of the system. However, at present no algorithm exists that can detect using low resolution images and at the same time deal with adverse lighting conditions. So in this thesis a light invariant, robust and efficient detection system is developed that can have manifold applications in the real world. These are summarized as follows:

- Surveillance Activities - Automated surveillance preventing unauthorized entry of humans, vehicles, and pedestrians in parking lots, traffic control systems or even in banks and airports where security is an important issue.
- Human-robot Interactions - Robotics and other such Artificially Intelligent applications often require continuous monitoring of human activities to detect changes in their behavior. Using motion to interpret changes is one of the simplest solutions in this case.
- Gaming Systems - In 3D video games motion sensing is often required to understand the movement of the player. This is usually deduced using embedded radio markers broadcasting the position of the player. With new automated detection systems this problem can be easily solved and the extra apparatus will not be required any more.

1.2 Objective

It is very easy for people to detect motion by observing changes caused in a stationary environment. However, it is a difficult problem to solve in computer vision because in reality the physical world is not an undisturbed stationary environment at all. In the case of people the ability to visually detect these changes in the environment is a natural process which is evidently not the case in a machine. Therefore the main objective in this thesis is to construct an automated detection system which can handle such non-stationary factors in real time. This system is based on a low level pixel by pixel approach and builds a background model initially by gathering information from the stationary environment.

The primary point of interest in this thesis is that it uses very low resolution images to build the model thereby introducing the concept of effective detection at a low price.

1.3 Research Scope

In this thesis, the problem of real time detection of moving objects using very low resolution images is addressed. This technique is then applied for selecting regions of interest from frames and storing that part in high resolution. This is actually achieved by simulating the behavior of the human visual system and proves the practicality of this work in real world applications. However this thesis does not resolve issues like dynamic updates of the background model and multiple level resolution approaches. These are left as future research topics. It should also be mentioned here that though different information sources as structure and color based techniques have been successfully used to detect the regions of motion they could be further extended to study human behavior in surveillance scenes. However that is beyond the scope of this thesis.

1.4 Approach

There are various techniques to solve the problem of motion detection. Vision scientists have approached this problem in multiple ways such as using Gaussian distribution to model a pixel or developing feature set vectors for extracting important information relevant to the process of detection. In the approach followed in this thesis an innovative technique based on low resolution input data is adopted to solve the problem. Instead of relying on a single source of information for building the system both structure and color based approaches are used to detect motion. Structure information is gathered from computing the Modified Census Transform on low resolution gradient images. Since the Modified Census Transform is invariant to light intensity changes, chance of misclassifying a pixel in varying lighting conditions as foreground when it is actually background or vice versa are signifi-

cantly diminished. In addition to the structure data obtained, color properties associated with every pixel in the form of temporal histograms are also used. These histograms are constructed without any prior information of the color distribution in the scene.

In Figure.1.1 the detection method is depicted which gathers spatial and color information from low resolution images corresponding to an input video. The focus is on detecting motion even in the presence of external environmental factors such as illumination fluctuations. The low resolution images used to train the system are initially converted to gradient images using Sobel operators.

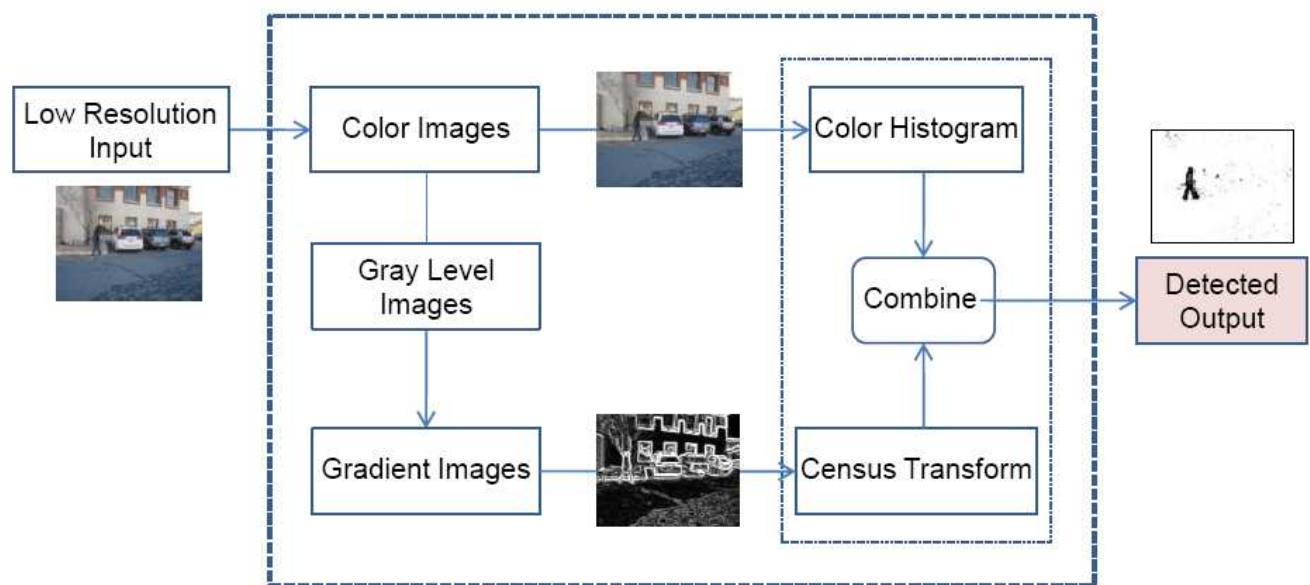


Figure 1.1: Flow of the Algorithm

1.5 Implementation

OpenCV forms the research platform in this thesis which includes library functions for real time computer vision. The library is mainly written in C, however, wrappers for C++ and Python are often used to make it more extensible. The structure information is stored using

a hash table data structure where the row entries correspond to pixels in every video frame and the column entries to values obtained from computing the Census Transform. In the case of color a double dimensional array is implemented to store the required information. While detecting in real time the values from the hash table and color array are combined in an effective fashion to identify the pixels that form the foreground.

1.6 Contribution

The most important contribution of this thesis is to demonstrate the possibility of reasonable real time motion detection for video surveillance. It depends upon the combined structure and color based approaches to identify the regions of interest. The second contribution is that it opens up new directions in low resolution imaging for detecting motion in images that have been extracted from a regular video sequence. Such cases are only possible because this thesis introduces a new technique to the engineering motion detection problem based on collecting information from various sources. The thesis also includes an interpretation of low level features to solve this problem. The third contribution is demonstrates that low resolution detection can also be an effective framework for developing other surveillance based systems as in foveation applications. Finally, it provides to the vision community an new automated real time low resolution robust detection tool, that works in adverse lighting conditions.

1.7 Outline

The rest of this thesis is structured as follows: Chapter 2 provides a brief overview of the related work in the field of motion detection, explaining its important concepts and features. In Chapter 3 that includes the main body of the thesis a detailed description of the detection algorithm based on using different sources of information is provided. Chapter 4 demonstrates the practicality and effectiveness of the approach by applying it to several

video based test cases. The idea is to view the classification of pixels belonging to a part of the stationary background or non-stationary foreground and then identify the region of motion associated with it. In the following Chapter 5 an introduction to the topic of foveation is provided. It also elaborates on motion based selection of regions of interest along with some preliminary results generated as first steps for developing a foveal visual system. Chapter 6 is a conclusion to this thesis with a summary of the findings, which is followed by an outline of future research directions.

Chapter 2

Previous Work on Motion Detection

Motion Detection algorithms are mainly divided into pixel-based methods and region-based methods. In pixel based methods, temporal differencing between frames based on binary thresholding is applied to identify the pixels that belong to the foreground. In the case of region identification entire zones are classified together by applying structure, color, or texture based methods. The different strategies that exist for performing real time robust motion detection will be summarized in the following sections.

2.1 Temporal Differencing

In this method individual frames are taken into consideration. Here the difference between consecutive frames from a single video sequence is computed to find out which pixels have changed and fall in the category of foreground. One of the consistent problems of this simple method is that often it is impossible to detect if there has been a change inside detected boundaries of motion. Also, in some cases pixel connectivities are not considered that limits the process of detection. Figure.2.1 shows an example where the differences between consecutive frames is computed and then thresholding is applied to detect the zone of motion. The white blob indicates the detected region of motion.

Such a differencing algorithm was implemented by Huwer et al. [HN00] where a temporal difference method was combined with an adaptive background model subtraction scheme to deal with lighting changes. Even though this approach offers somewhat satisfactory results, much research is still needed to solve the motion detection problem, and thus complementary and alternate approaches are worth investigating. Figure.2.2 shows an

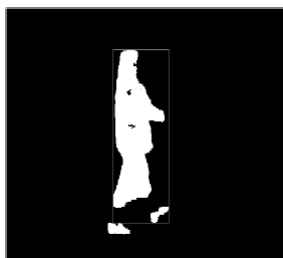


Figure 2.1: Temporal Differencing e.g.1

example where there has not been much motion between two consecutive frames and as a result of which there has only been a slight change in the edges signifying movement.



Figure 2.2: Temporal Differencing e.g.2

2.2 Background Modeling

In background based modeling, the traditional approach for detecting motion consists of building a model of the background using multiple frames and then classifying each pixel in the surveillance frame as either foreground or background. The average of all images across a video sequence is used to generate the background model and later the pixel likelihoods

are computed by comparing with the average. In cases where the value is significantly different the pixel is classified as a foreground pixel. Some such existing approaches include Ren and Chua [RC01], who propose spatial distribution of Gaussian(SDG) models, where motion compensation is only approximately extracted, Stauffer and Grimson [SG99], who model each pixel as a mixture of Gaussians and uses an online approximation to update the model and Monnet and Mittal [MMPR03], where an online method based on dynamic scene modeling is used.

Wren et al.[WADP97] also designed a system for detecting and tracking the human body. Here unlike the previous case the background model used a Gaussian distribution to update the status of every pixel in the frame. The person is represented using spatial and color information and then each time the person undergoes a change the spatial information gets upgraded using a Kalman filter. In the next step a probabilistic model of estimation decides whether each image pixel is a part of the background or the person. The detected pixel is then assigned to the corresponding zone it belongs to and the background model is again updated. The spatial information is obtained from different sources to determine the objects contour whereas color information can be determined as a function of the color of the person being detected.

More recently, some hybrid change detectors have been developed that combine temporal difference imaging and adaptive background estimation to detect regions of change [HN00]. Krum et al [JKM] designed a model where a three tier algorithm is followed and processing is done at the pixel, region and frame levels. However, in 2004, Beleznai et al.[BFB04] came up with the idea of using difference in intensity between the actual frame and a reference frame to represent the system with a probability distribution model. Motion was detected using mean shift and the concept of integral images was used that made the model highly efficient by increasing the computational speed. When detection of clusters was required some modifications to the original model was used.

2.3 Optical Flow

Optical flow models are based on determining the velocity of pixels in an image in order to identify the objects of motion. Some of the works where this technique has been used include gradient methods where differential derivative of components are computed, sub-image window method where shifts in windows for every frame is detected to calculate displacement for the central window pixel. Even then not much motion can be estimated because of the approximation in the constraint equation. The most well known optical flow algorithm was introduced by Lucas and Kanade [LK81] based on frame differencing. It assumes constant flow and takes into account local spatial neighborhood around a given pixel at any point of time. This algorithm uses combined information from different methods like statistical methods and stochastic models. In models for detecting motion or stereo matching techniques the algorithm works in a step down approach.

An optical flow application was developed by Elzein et al. [HSP03] in 2003 where it was used on specific regions of the frame obtained from temporal differencing. The camera was mounted on a vehicle and the time for collision with an object was then estimated with respect to a reference point using velocity of flow. The objective was to detect objects which suffer a collision with the camera. This was computed using connected components of pixels. Later this object was segmented to be classified as a pedestrian or a non-pedestrian.

2.4 Entropy Based Models

These models assume that a pixel's state would undergo a small change if disturbed by noise and a larger change if disturbed due to motion. Such a state change can indicate whether the pixel is undergoing any change due to motion or not. The idea is to use accumulating windows to define the relationship between the pixels concerned and use color variations to identify the entropy of the image as an energy representation. Here the higher the level of energy the higher are the chances of that pixel being classified as undergoing motion. One

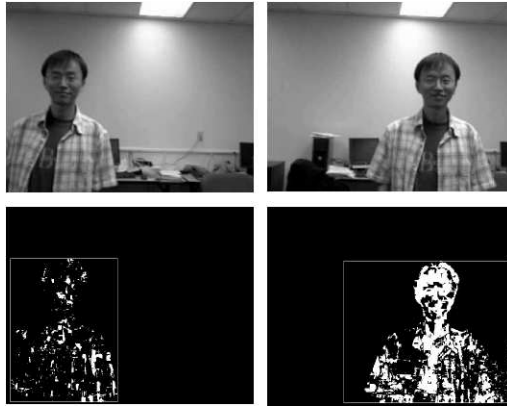


Figure 2.3: Detection using Optical Flow Algorithm

of the main advantages of this method is that the computational cost is much lower than other methods for detecting motion. However, all motion capturing devices work based on the simple principle that motion objects are moving with respect to stationary background. Since it is often impossible to remove all noise sources the accuracy obtained as a result is much lower. Several techniques have been applied to prevent the occurrence of noise but none are good enough to remove all sources.

2.5 Direct detection

In addition to the already stated methods there exist others that are based on their mixture in order to solve the problem of motion detection. One such example is the work of Cutler and Davis [CD00] that focuses on detecting periodic motion and then applying this periodic motion to identify patterns in biological subjects. Here initially both frame differencing and thresholding are used to detect regions of motion and then other operations are used to obtain the set of objects that are being tracked. For detection and analysis Fourier Transform coupled with correlation is used. To distinguish between different biological forms such as humans and non-humans the model is slightly modified. The performance of

the system is in real time.

In 2002, Utsumi and Tetsutani [UT02] used similar methods to detect motion in different forms using the fact that geometric distances between various body parts of humans are the same even though there are variations in the pixel intensities due to changes in color and lighting conditions. Using a distance matrix to represent the distance in color distributions they built a model corresponding to distance matrices of both humans and non humans. The different models are then compared and found out to be quite similar in nature. So using the slight variations in the computed models they could classify an image to be human or non human. Another well known work worth mentioning here is that of Viola and Jones [VJS03] who performed direct detection from images using Haar features and classified using an ensemble of classifiers. The learning algorithm used was Adaboost and detection was at a fast and appropriate rate.

Chapter 3

Solving Motion Detection

In order to achieve robust detection, a combination of information from different sources, namely structure and color is used. The structure information is computed using the Census Transform and color information is based on generating Temporal Color Histograms. Both of these techniques are briefly described later in this chapter. Figure.3.1 demonstrates the working of the algorithm using these techniques. The Census Transform actually determines information about the spatial organization of the object that is detected whereas color histograms essentially stores information with respect to the color distribution in RGB space. Once the values corresponding to these information sources have been successfully computed for every pixel in the frames a look up table is constructed to map their

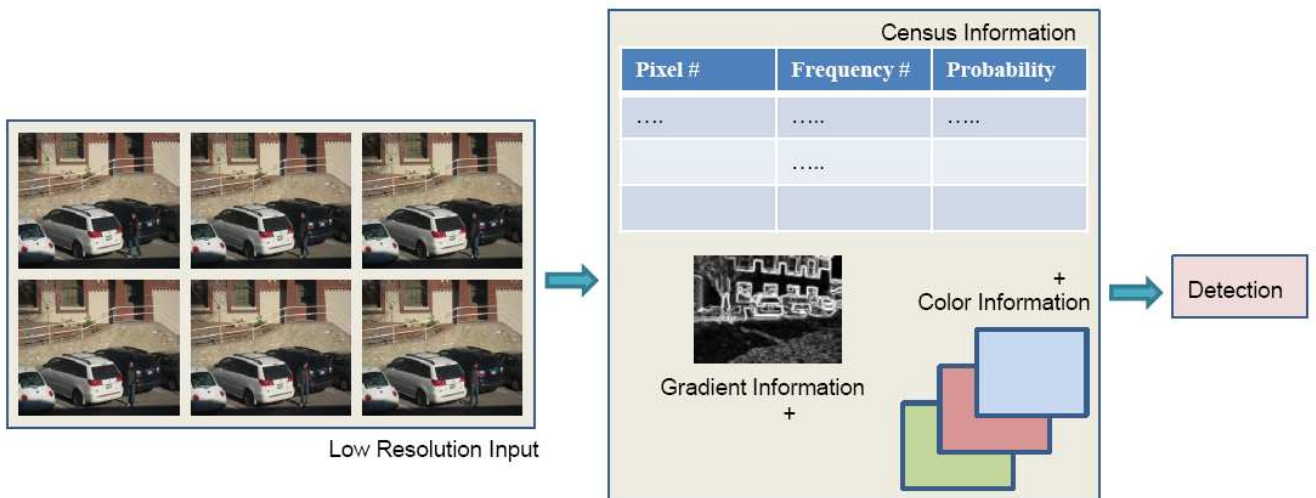


Figure 3.1: Working Principle

3.1 Introduction

3.1.1 The Census Transform

The Census Transform is a non-parametric summary of local spatial structure. It was originally proposed in [ZW94] in the context of stereo-matching and later extended and applied to face detection in [FE04]. It has also been used for optical flow estimation [Ste04], motion segmentation [YMAS01], and hand posture classification and recognition under varying illumination conditions [JRS06]. The main features of this transform are also known as structure kernels which are used to detect whether or not a pixel falls under an edge. The structure kernels used in this paper are of size 3×3 , however, kernels can be of any size $m \times n$. The kernel values are usually stored in binary string format and later converted to decimal values that denote the actual value of the Census Transform.

In order to formalize the above concept, let us define a local spatial neighborhood of the pixel x as $N(x)$, with $x \notin N(x)$. The Census Transform then generates a bit string representing which pixels in $N(x)$ have an intensity lower than $I(x)$. The formal definition of the process is as follows: Let a comparison function $\zeta(I(x), I(x'))$ be 1 if $I(x) < I(x')$ and 0 otherwise, let \otimes denote the concatenation operation, then the Census Transform at x is defined as $C(x) = \otimes \zeta(I(x), I(y))$. This process is shown graphically in the Figure 3.2.

3.1.2 The Modified Census Transform

The Modified Census Transform (MCT) [FE04] is an extension of the Census Transform first introduced by Zabih and Woodfill [ZW94]. Let (r, c) be a pixel position in image I . Let $N(r, c)$ be an ordered set containing the pixels in the 3×3 neighborhood of pixel (r, c) in I . The MCT generates a string of nine bits representing which pixels in $N(r, c)$ have an intensity that is greater than the average intensity in $N(r, c)$.

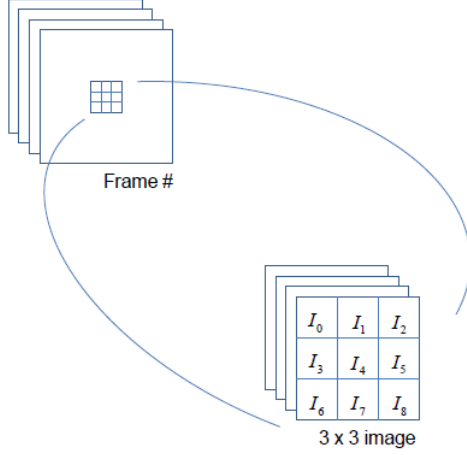


Figure 3.2: Census Transform

$$N(r, c) = \{(r', c') | r' \in \{r - 1, r, r + 1\}, c' \in \{c - 1, c, c + 1\}\} \quad (3.1)$$

$$\mu_I(N(r, c)) = \frac{\sum_{(r', c') \in N(r, c)} I(r', c')}{9} \quad (3.2)$$

$$MCT_I(r, c) = \bigoplus_{(r', c') \in N(r, c)} \xi(I(r', c'), \mu_I(N(r, c))) \quad (3.3)$$

where \bigoplus denotes concatenation and ξ is an indicator function such that:

$$\xi(x, y) = \begin{cases} 0 & \text{if } x \leq y \\ 1 & \text{otherwise.} \end{cases} \quad (3.4)$$

3.1.3 Temporal Color Histogram

The color histogram is a compact representation of color information for every pixel in the frame. These are flexible constructs that can be built from images in various color

spaces, such as the R-G-B, H-S-V and L-a-b. A histogram of an image is produced first by discretization of the colors in the image into a number of bins, and then counting the number of image pixels in each bin. Color Histograms are used instead of any other sort of color cluster description [HKR97, JC98, MRG99, NO99], due to their simplicity, versatility and velocity, that are needed in tracking applications. Moreover, it has been vastly proven its use in color object recognition by *color indexing* [FF95, SB91]. However, the major shortcoming of detection using color is that it does not respond very well to changes in illumination and noise artifacts. Figure.3.3 shows such a color space model where each of the faces of the cube represent a different color along each of the axes.

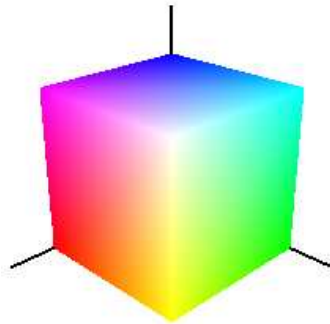


Figure 3.3: Color Cube Model

3.2 Proposed Algorithm for Detection

The proposed algorithm mainly consists of two parts; the training procedure and the testing procedure. In the training procedure a look up table is constructed using the background information. Here we consider individual pixels in background video frames and then identify their gray level and color values in the R-G-B space. The structure kernels and

color cubes are then computed using the Census Transform and Color Histograms based on these values. These later serve as indices in the look up table that is constructed and the contents at those indices are counters for the number of frames that have the same structure kernel and color cube number. The look up table is next normalized to assign probability values at every location. In the testing phase of the algorithm these probabilities are retrieved from the look up table and combined together to identify the pixels that demonstrate motion. The flow chart diagram in Figure.3.4 briefly summarizes all the sequential activities that are performed in order to do the detection.

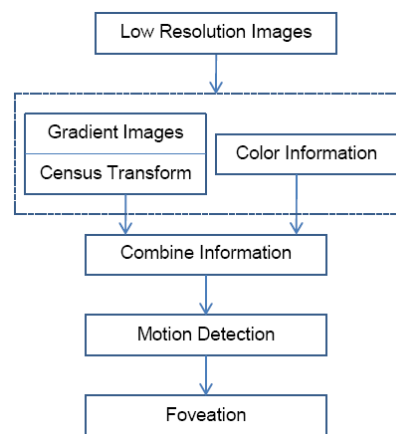


Figure 3.4: Flow Chart

3.2.1 Training

Ternary Modified Census Transform

The modified Census Transform generates structure kernels in the 3×3 neighborhood but the kernel values are based on slightly different conditions. So it has been found to be useful in the detection of high-contrast objects, such as faces, however, when applied to regions of uniform color it is very sensitive to noise, which makes it unreliable for motion detection.

To solve this problem, we proposed a base-3 census transform, where each element of the descriptor indicates whether a pixel’s intensity is significantly higher, significantly lower, or approximately equal to its neighborhood average. Thus the base-3 MCT is given by:

$$\xi = \begin{cases} 0 & \text{if } x < y - \lambda \\ 1 & \text{if } x > y + \lambda \\ 2 & \text{otherwise} \end{cases} \quad (3.5)$$

where λ is a positive constant that acts as a threshold. Once the value of the Census Transform kernel for a pixel has been computed it is used to index the occurrence of that pixel in the look up. So for every pixel in the background frames kernel values that form the indices of the look up table are generated. For any index in the look up the contents are the frequencies of frames that have the same value of the structure kernel. Once the entire lookup has been constructed from the background video the probabilities are computed that are then used directly at the time of testing.

Color Histograms

The color probabilities are obtained in a similar fashion but using the concept of color histograms. In this case initially the color intensities are obtained in R-G-B format and based on these values the color cube/bin corresponding to the Red-Green-Blue values of the pixel is determined. Since each color is quantized to 4 levels a sum total of $4^3 = 64$ such bins are possible. Every pixel then has a corresponding counter containing 64 values and the color cube values form the indices of the counter look up table. The contents of the table are the number of frames that have the same counter value for that pixel.

3.2.2 Testing

The frames from the test video are used to obtain the pixel intensities in both gray level and R-G-B color space. Then for each pixel we recompute the Census Transform. This serves as an index in the existing look up table for the probability corresponding to that

Algorithm 1 Training:Structure Information

Require: Images extracted from background video sequence.

Ensure: Build background model B_g .

- 1: **for all** $frame\#0$ to $frame\#N$ **do**
- 2: **for all** $pixel\#(0,0)$ to $pixel\#(m,n)$ **do**
- 3: Compute $N(r,c)$ using Equation3.1
- 4: Compute the value of $\mu_I(N(r,c))$ based on $N(r,c)$ using Equation3.2
- 5: Compute $\xi(x,y)$ using Equation3.5
- 6: Compute $MCT_I(r,c)$ from $\xi(x,y)$ using Equation3.3
- 7: Increment $frequency\#$ entry in look up table where $MCT_I(r,c)$ serves as index
- 8: **end for**
- 9: **end for**
- 10: Normalize look up table by changing from $frequency\#$ to $prob\#$

NOTATION:

$N(r,c)$ is an ordered set of pixels in 3×3 neighborhood of pixel (r,c) in I

\oplus denotes concatenation and ξ is an indicator function

pixel. In this way we use the probabilities corresponding to all pixels in any frame being processed in real time. This procedure is depicted in Fig.3.5 that demonstrates the concept of testing where initially the Modified Census Transform is computed for pixel_{*i*} based on its 3×3 neighborhood. The kernel value is used as an index for the look up table already constructed to retrieve the required frequency. Here the value of the Census Transform is 25 that then serves as the index. The frequency 33 actually represents that for pixel_{*i*} there are 33 frames that have the same kernel value of 25.

In the case of color the same scheme is used where every pixel is split into its R-G-B color intensities and then its color bin is computed. Then the probability corresponding to color bin computed is retrieved from the look up table for that pixel. This generates structure and color probabilities corresponding to every pixel in the frame. These two probability values are then combined in an efficient way to detect the zones of active motion.

Algorithm 2 Training:Color Information

Require: Images extracted from background video sequence.

Ensure: Build background model B_g .

- 1: **for all** $frame\#0$ to $frame\#N$ **do**
 - 2: **for all** $pixel\#(0,0)$ to $pixel\#(m,n)$ **do**
 - 3: $pixel\#(r,c)$ in $frame\#I$ is split into corresponding $R-G-B$ values
 - 4: Identify $colorcube\#$ corresponding to that pixel based on its $R-G-B$
 - 5: Increment $frequency\#$ entry in look up table where $colorcube\#$ serves as index
 - 6: **end for**
 - 7: **end for**
 - 8: Normalize look up table by changing from $frequency\#$ to $probC\#$
-

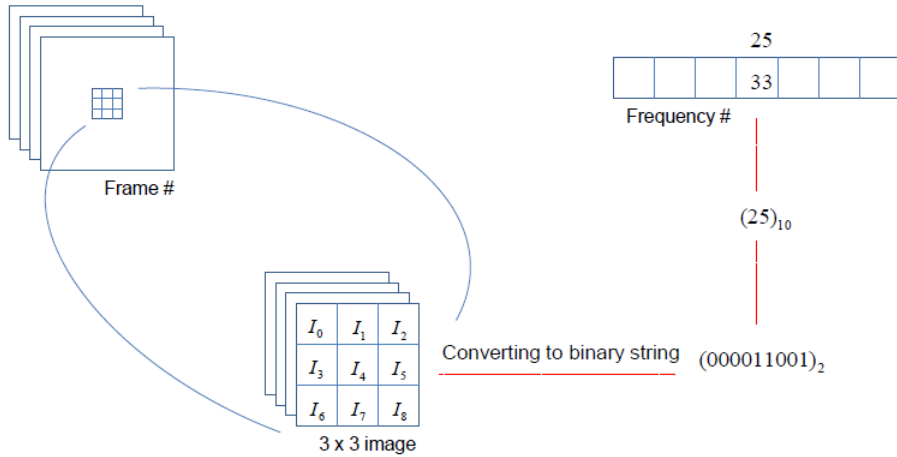


Figure 3.5: Testing Algorithm

3.2.3 Combining Information from Color and Structure

Once the corresponding structure and color probabilities are obtained for every pixel of the test frame the lower probability value is selected for thresholding that pixel. Based on the value of threshold in the experiment the pixel is classified as background if the probability

is more than the threshold and as foreground if the probability is less than the value of the threshold. The algorithm for the testing and classifying each pixel has been represented as follows:

Algorithm 3 Testing:Structure and Color Information

Require: Images extracted from foreground video sequence.

Ensure: Classify pixels in foreground frames

- 1: **for all** $frame\#0$ to $frame\#N$ **do**
 - 2: **for all** $pixel\#(0,0)$ to $pixel\#(m,n)$ **do**
 - 3: Compute structure information using Modified Census Transform
 - 4: Compute color information using Temporal Color Histograms
 - 5: Find probability from B_g given structure information
 - 6: Find probability from B_g given color information
 - 7: Combine probabilities
 - 8: Classify $pixel\#(r,c)$ based on threshold
 - 9: **end for**
 - 10: **end for**
-

Chapter 4

Results and Observations

4.1 Experiments Conducted

The first set of experiments were conducted in an outdoor environment where there were variations in light intensity. The training set is a background video consisting of 900 frames recorded in raw format at a rate of 30 frames/sec and at a resolution of 640×480 pixels. The detection method is based on initial computation of the gradient images from the video and then determining the Census Transform information from the already computed gradient images. Each pixel has a corresponding Census value for all the 900 frames and the frequency of a particular Census value for that pixel is stored in the look up during the training phase of the algorithm. Similar scenario happens for constructing the look up of the color values where the RGB color space corresponding to every pixel is considered.

4.2 Detected Results

The testing phase of the algorithm starts after the background model using all the 900 frames has been built. Here for every pixel in the test frame the combined probability values are looked up in real time and displayed accordingly. As a result a pixel with low probability in the test video has a corresponding low or zero probability assigned to it. In this case the test video has 1001 number of frames that get detected in real time and the resultant image displays the the detected zone of motion. The frames in the test video are also stored at a rate of 30 frames per sec and a bit rate of 705kbps. In Figure.4.1 the frames are displayed at a resolution of 640×480 pixels whereas in Figure.4.2 the corresponding

binary images depicting the stage after detection are displayed at a much lower resolution. Here detection was carried out on a training set with exactly half of the original 640×480 resolution. Similarly the resolution is decreased to one fourth of the original for the training set used in Figure.4.3. The corresponding low resolution and detected images are both displayed in Figure.4.3 and Figure.4.4.

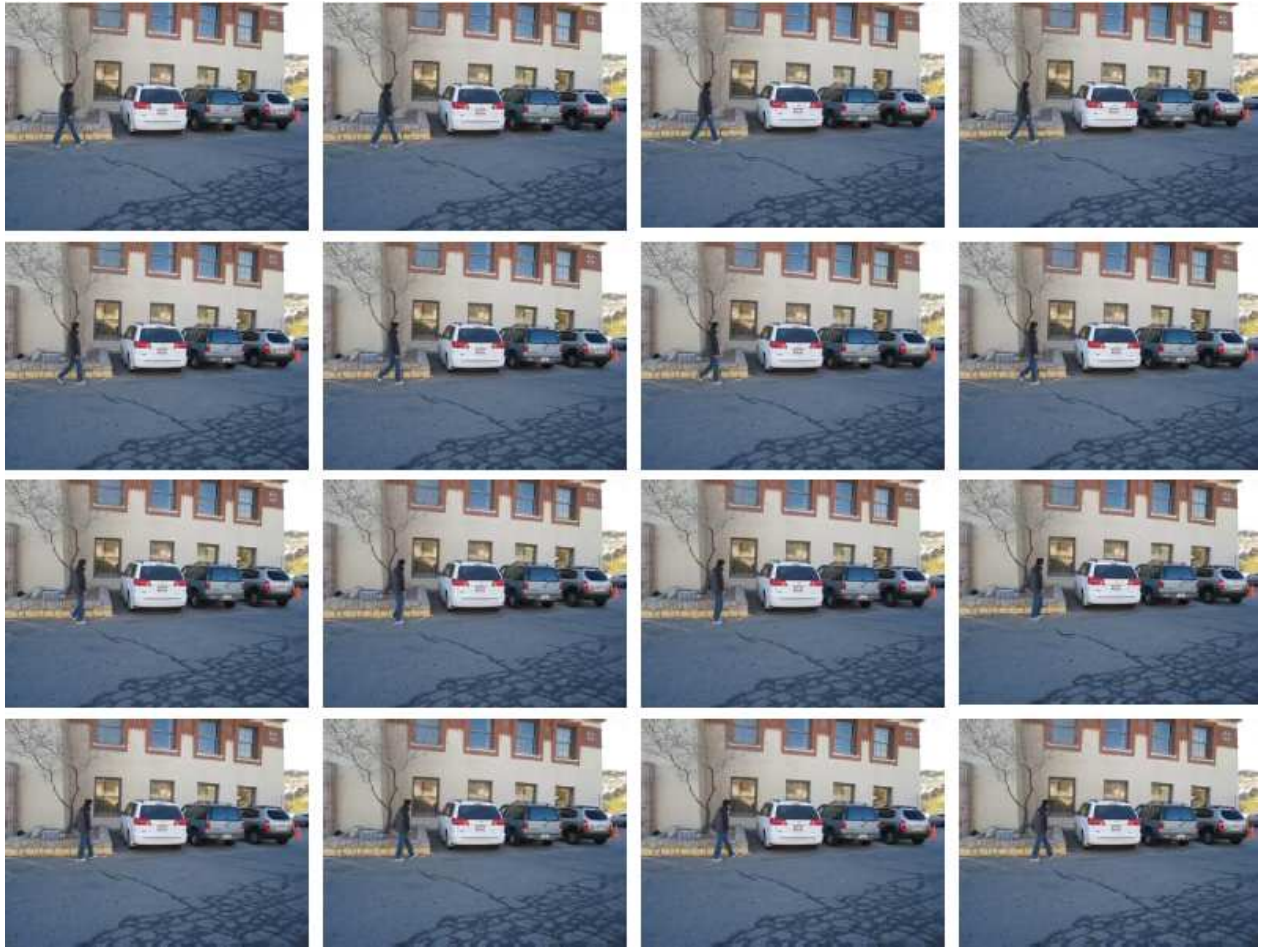


Figure 4.1: Frames 70-86 from OutdoorScene1

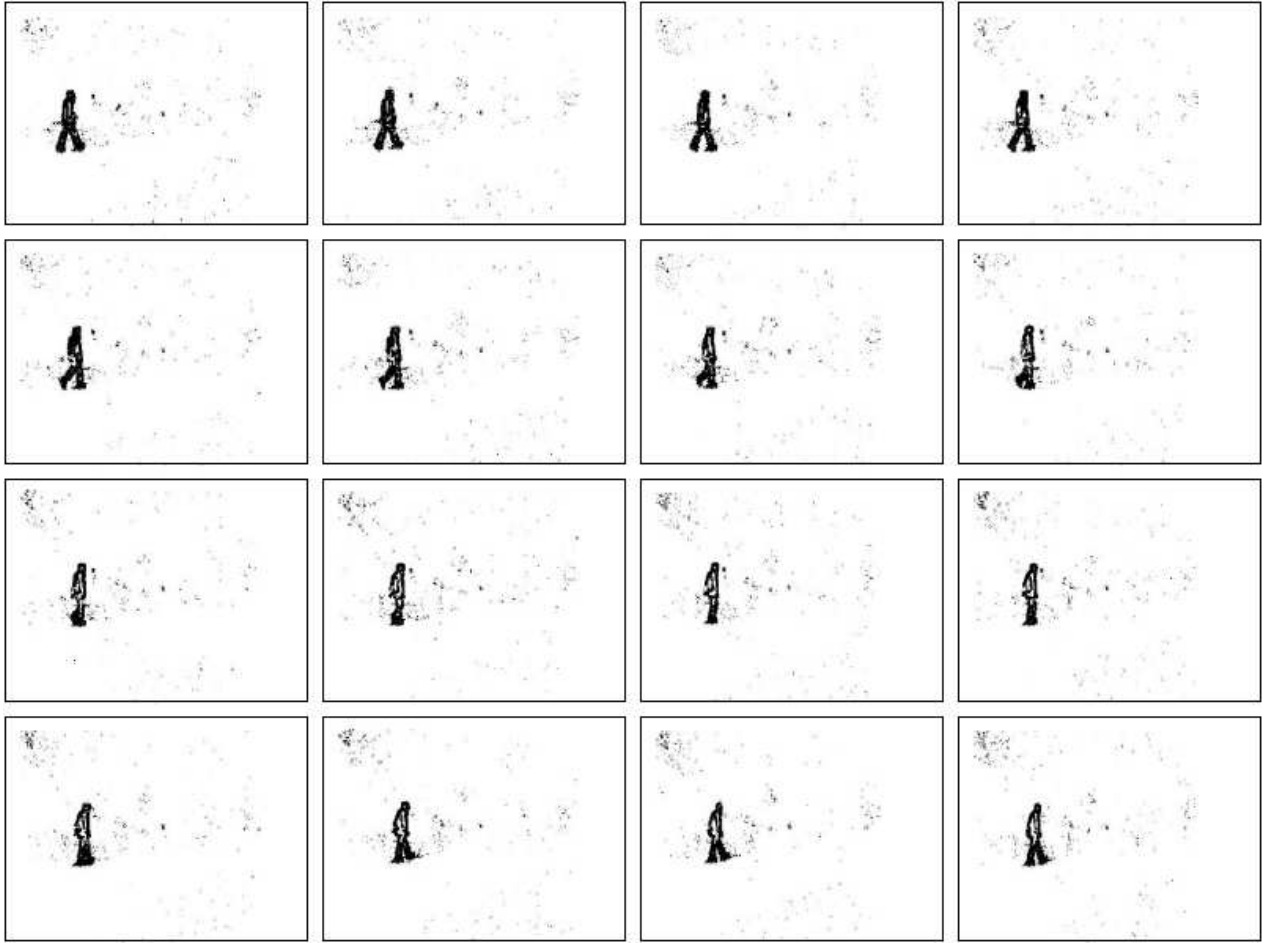


Figure 4.2: Detected Frames 70-86 from OutdoorScene1

4.3 Analysis of Experiments

In order to quantitatively evaluate the results of the method, each pixel is manually classified in the sequence as either foreground or background and then this classification is compared with that provided by the algorithm. In Figure 4.5 and 4.6 the ROC curves corresponding to four different resolution levels are depicted.



Figure 4.3: Frames 50-58 from OutdoorScene1

4.3.1 Sensitivity to Low Resolution

This subsection examines the relationship between accuracy and resolution. Recall that in Figures 4.5 and 4.6 the ROC curves for different resolution levels are depicted for two outdoor scenes. Performance at 1/4th resolution is inferior to 1/16th resolution. The higher performance at 1/16th resolution may be due to noise reduction due to averaging. Performance at sub-sampling beyond 1/16th provides inferior performance – possibly due to the large aggregate pixel size. This proves that decreasing resolution provides a level of

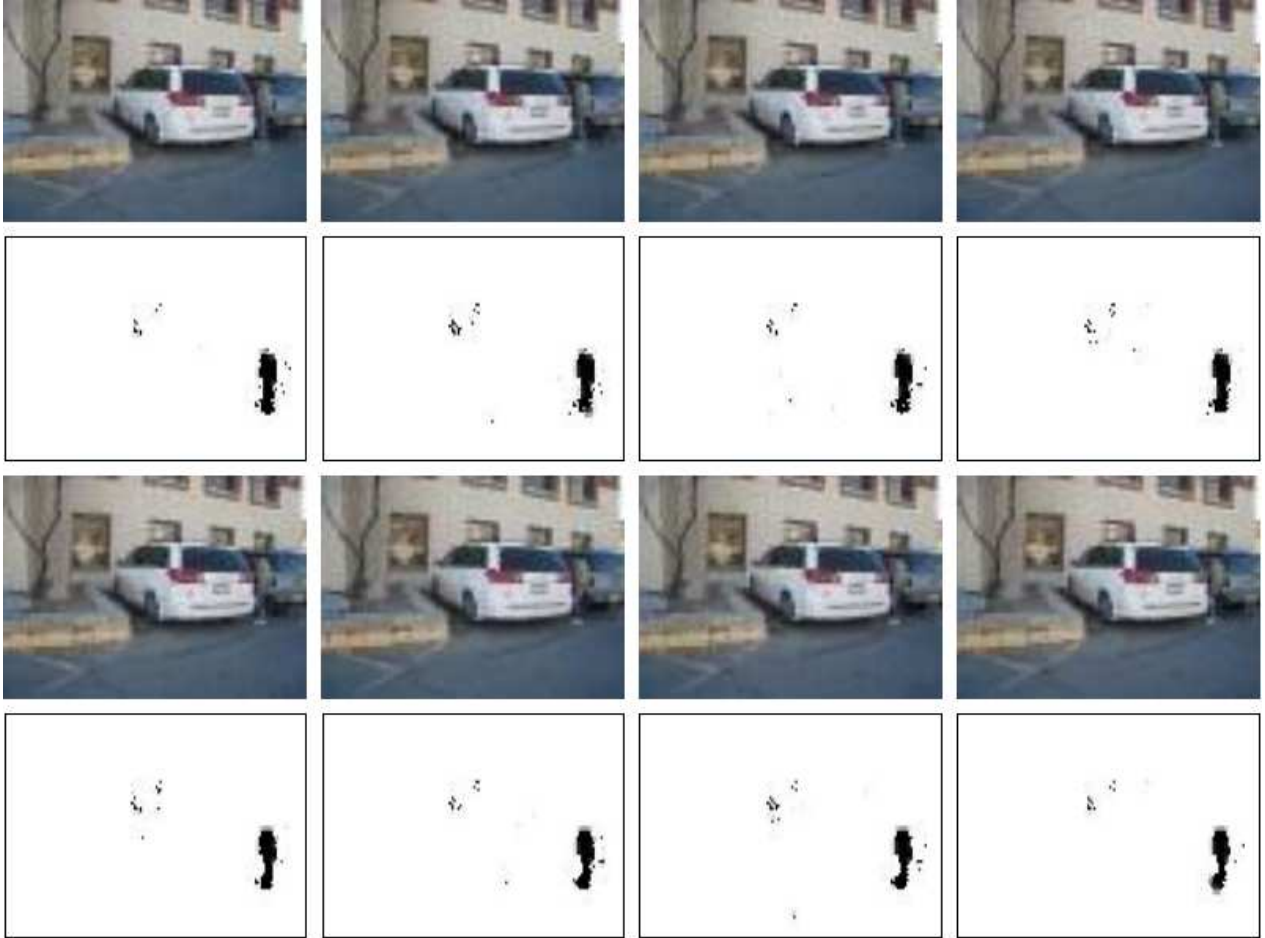


Figure 4.4: Frames 315-323 from OutdoorScene2

noise-filtering by means of averaging pixel values, but at very low resolutions identifying small regions of motions becomes more difficult.

Another observation made from the experiments conducted was the increase in efficiency using the low resolution approach. The execution time in milliseconds for both experiments is tabulated in Table4.1. This is required to perform training for the dataset corresponding to the background videos OutdoorScene1 and OutdoorScene2. From the results obtained in the two experiments it can be easily seen that the times are almost similar with only small differences of not more than 60 milliseconds. It can also be deduced from

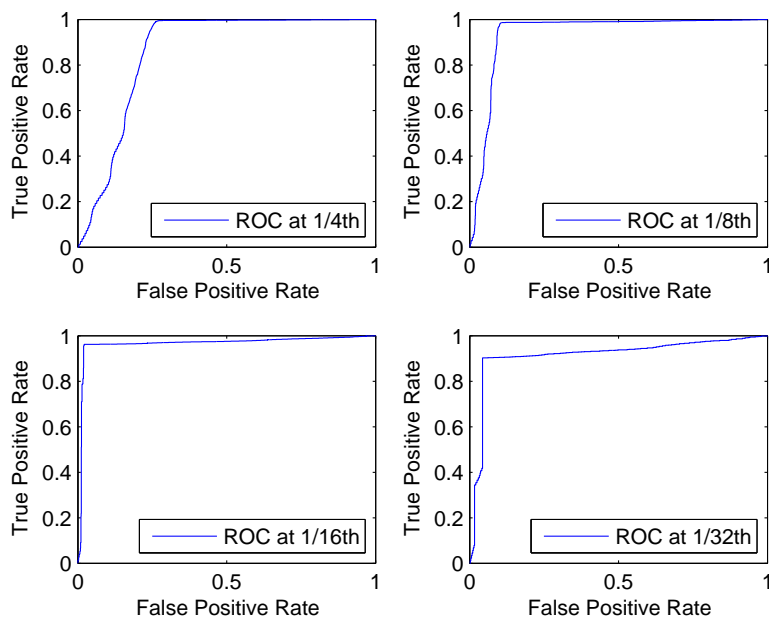


Figure 4.5: ROC curves for OutdoorScene1

the different instances that the gradual lowering of the resolution resulted in the decrease in computational time for the training datasets.

Table 4.1: Different Training Times(ms)

Resolution	Time1	Time2
1/4th	3125ms	3078ms
1/8th	3066ms	3063ms
1/16th	3041ms	3047ms
1/32th	2953ms	2969ms

Figure 4.7 illustrates a comparative analysis of ROC curves in order to understand the variations in the nature of the curves for different test scenarios. Training Set1 is based on OutdoorScene3 where there are more diverse changes in the weather conditions such as wind fluctuations, lighting changes etc. whereas Training Set 2 is based on OutdoorScene1. In the curves it is distinctly observable that the accuracy in classification has a significant

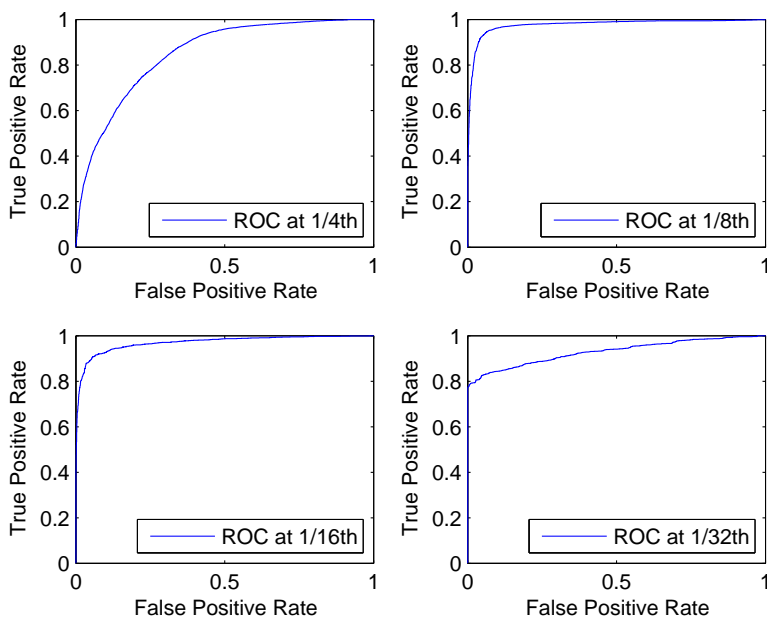


Figure 4.6: ROC curves for OutdoorScene2

drop when the resolution is reduced to $1/4$ of the original. However, Training Set 2 performs much better than Training Set 1 because of the mentioned outdoor parameter fluctuations. But as predicted already in the above testing scenarios the intermediate resolutions achieve better results due to the almost redundant noise factor.

4.4 Observations

An extensive evaluation of the detection system built involves handling multi objects and these objects cross each other in sequence of time. However, the detector system is robust enough to handle such cases that involves multiple object detection. Also if the object disappears behind another object that formed part of the background then the system would stop detecting the object as desired. Another difficult case of detection would be in cases where two objects merge as a single object or one single object splits into two halves. But since the method is based on background modeling using feature extraction

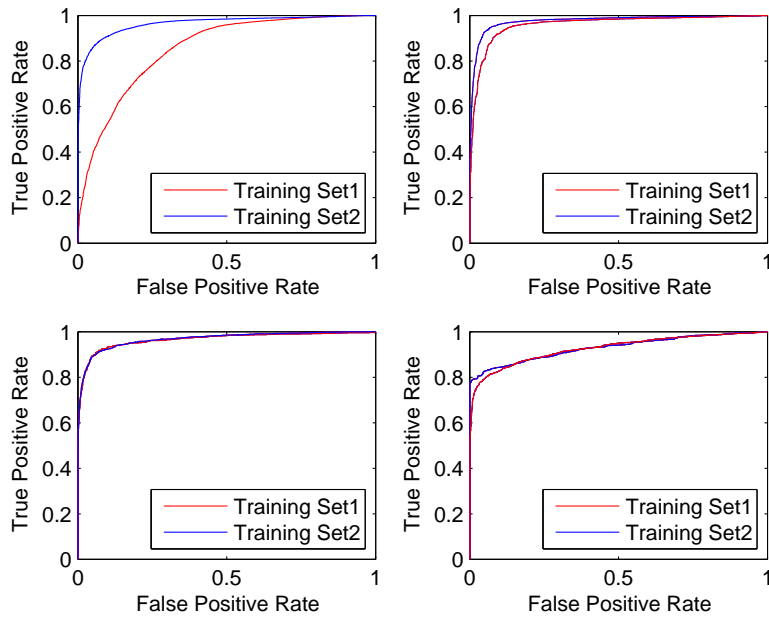


Figure 4.7: Comparison of ROC curves

any changes in the background would be detected as the foreground even if it involves multiple objects. One thing has to be kept in mind however, that since the input to the system is very low resolution images the detection is also maintained at the same level. So often some pixels get misclassified as detected pixels. That increases the number of false detections. A fair estimate of the percentage of misclassifications can be formed by quantitatively analyzing the number of true positives and false positives from the curves.

Chapter 5

Foveation

5.1 Introduction

Traditional approaches to video surveillance transmit full-frame video imagery to centralized analysis centers. Unfortunately, the large amounts of bandwidth and power required for detection and transmission of high resolution full-field video, limit its use in deployments for pervasive surveillance. Foveation, where specified video scene areas of high interest are transmitted in higher resolution than the surrounding scene, is useful for reducing the bandwidth required for high resolution surveillance.

Foveation, is inspired by the non-uniform resolution perception of the human visual system due to the non-uniform density of photoreceptor cells in the eye and offers a promising way to deal with these problems. While traditional approaches to video surveillance utilize full-frame imagery, which is either analyzed locally or transmitted to remote analysis centers, a foveated system can focus on providing high-resolution only in task-relevant regions of interest in the image, thus dramatically reducing power and bandwidth requirements and preventing data overload.

Inspired by the efficiency of foveation in biological systems, computer vision researchers have built two types of working foveal vision prototypes. One approach consists of building non-uniform resolution in hardware, normally trying to emulate the distribution of light-sensing elements in the human eye [TS93]. The main restriction in these systems is that the foveal area is confined to a single fixed area in the image plane, normally the center. Thus, taking advantage of foveation requires mounting the camera on a pan-tilt head, which makes its operation slow and cumbersome and generates additional power requirements.

The other approach consists of using a conventional high-resolution camera and transferring full-frame images to local memories that are sub-sampled in various ways prior to transmission [MA06]. While a tiny fraction of sampled pixels are transmitted, the sensor and its controller must nonetheless process full untransmitted frames, limiting savings in power consumption. Moreover, the selection of regions of interest has not been addressed in these works, thus it is not clear how they can be used to limit data production.

New sensors are currently being developed that will be capable of sensing of multiple foveal areas located at arbitrary positions within an image. This will allow the exploitation of many of the potential advantages of foveation.

A challenging problem lies in determining, from the original low-resolution image, the location of the regions of interest where increased resolution is required. In video surveillance, motion is very useful to identify potential regions of interest. In the previous chapters of this thesis an algorithm for robust motion detection for low resolution images is presented using cues from structure and color. In this chapter preliminary results to develop a system that uses the detector algorithm to identify and foveate the region of motion is presented. The foveal system is specifically designed to work reliably at low resolutions.

5.2 Some Related Work on Foveal Techniques

Foveal vision has been used for multiple purposes some of which have been approached by known techniques such as the Log Polar Transform, Foveal Wavelet Transform etc. Filtering is a technique where an image of uniform resolution is foveated in such a way that the region of interest is determined by imitating the behavior of the human eye. The log polar transform keeps the Cartesian properties of the image intact as shown by Martinez et al [MA06] in which a new approach for object tracking is presented.

A different kind of system adopted by several researchers is to compress the image data especially in video formats where due to elimination of some pixels bordering the periphery the whole image can be efficiently stored. One of the common techniques include

the logmap method [WOBS94] where the concept of space variant image processing is used assuming human vision to be highly space variant. Also a connectivity graph generates the behavior of the neighbors in such image variant random sensors. In other cases such as the multistage pixel method used by Kortum et al. [KG96] the resolution of the human eye is also assumed to be space variant and their algorithm is based on space variant degradation techniques. Here eye movements are also taken into account to provide the correct alignment for high resolution part of the image to that of human vision.

Bandera and Scott [BS89] also demonstrate such an approach using foveal sensors. Arrebola, Camacho and Hernandez [ACH97] on the other hand use a shifted foveal technique where the concept is centered around using different degrees of freedom along with higher number of fixation points from the shifting algorithm to reduce the error rate in positioning the fovea. In some others like [KGG99] [Kuy98] uniform grid techniques based on retinally reconstructed images are used. The retinal images are constructed depending on the codes generated from positions of sampling.

Geisler et al [GP98] uses multi resolution pyramid structures to code images into varying degrees of resolution. Lee, Pattichis and Bovik [LPB98], [Lee00], [LPB01], [LPB02] use low pass filtering to obtain the foveated images. The results are quite good because large chunks of redundant pixels at very high resolution have been removed. Some other existing methods include use of wavelet foveation methods like in [CY97], [cC98], [CMY99] where the multi resolution framework introduced by Mallat is used. Apart from their work on wavelet foveation Shapiro et al also used such techniques to develop the Embedded Zerotree Wavelet algorithm in [Sha93].

5.3 Foveal Detection

Once all the information from the Census Transform and the Color Histograms has been computed, the values are retrieved from the pixel look up tables for the test video. The training for the algorithm is carried out using entire sequences of low resolution frames

whose low level features computed using the structure and color information are stored for later retrieval. Next, the pixels from the test video frames are used to generate the kernel number for structural references and bin number for color information. These values are used to compute the likelihood of each pixel to be classified as background or foreground.

After a preliminary classification is obtained using the original algorithm, we examine the neighborhood of each pixel classified foreground (that is, presenting motion). If any of the pixels in the 3×3 neighborhood surrounding a pixel originally classified as foreground is not also a foreground pixel, then the classification of that pixel is changed to background. This eliminates isolated false detections due to noise and also removes spurious detections around the contours of moving objects

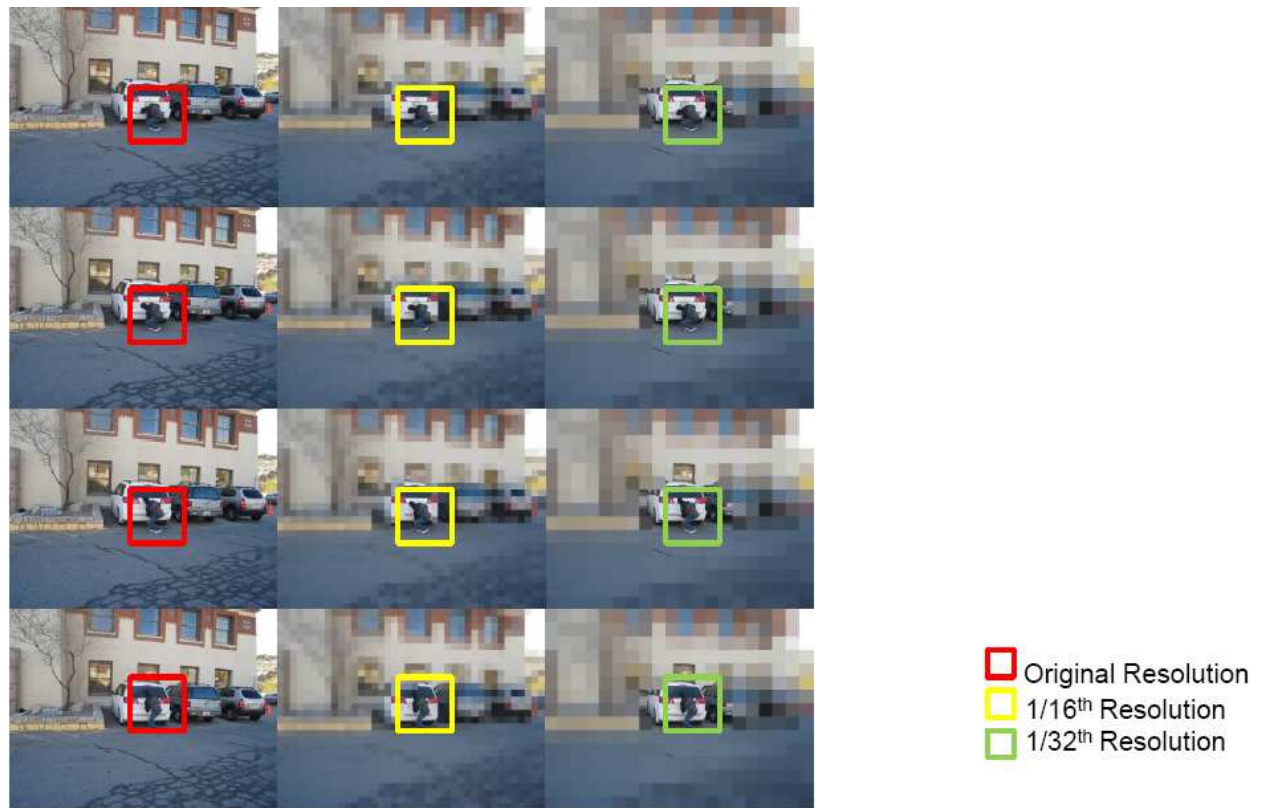


Figure 5.1: Foveation in OutdoorScene1

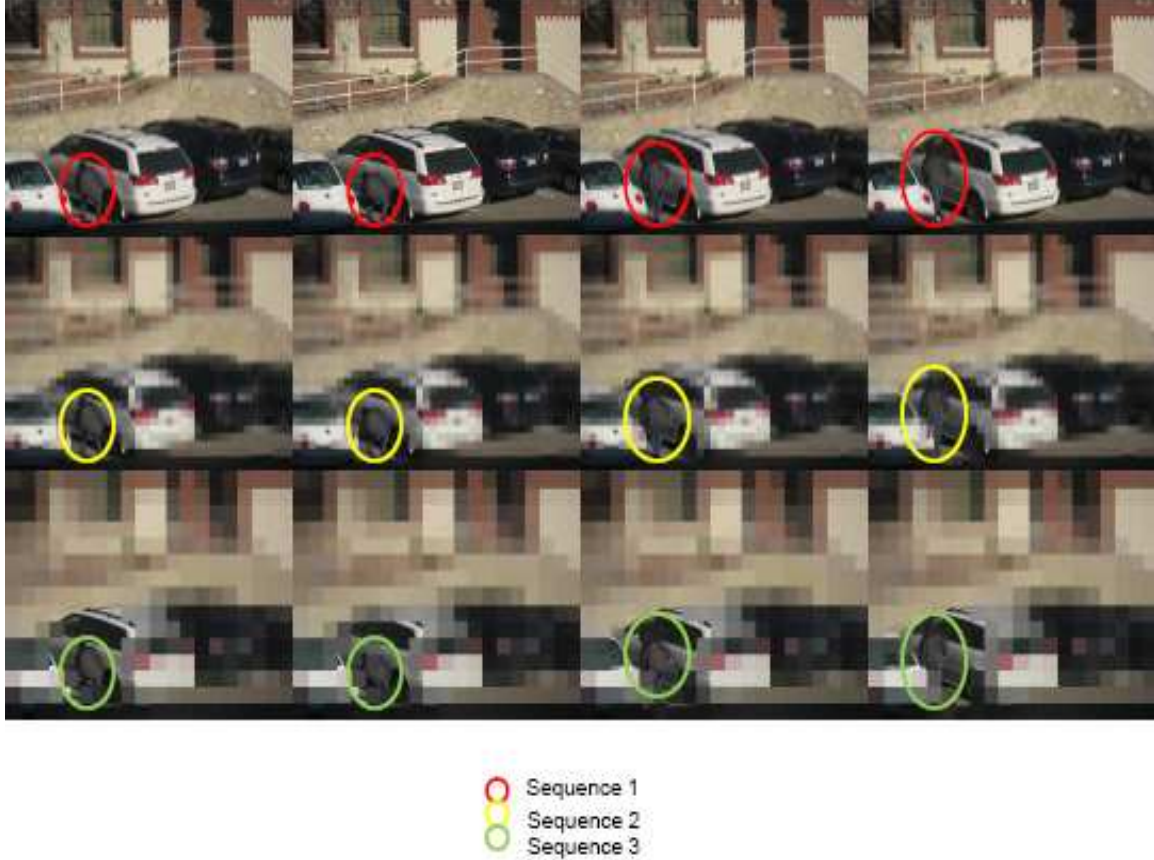


Figure 5.2: Foveation in OutdoorScene3

5.4 Experimental Results

A foveal camera is simulated by first capturing a video sequence in high resolution, obtaining a low resolution sequence from it, processing all information using low resolution and then foveating by replacing the selected low resolution pixels by the corresponding high resolution ones from the original image. In Figure 5.1 detection is carried out using a training data set of a video sequence which has 1680 frames in it. The video is recorded in an outdoors environment in a parking lot. The characteristics of the input video used are enumerated in table 5.1. Low resolution frames are generated using sampling techniques that serve as training data for the foveal system. The various levels of training are at 160×120 , 80×60 , 40×30 , 20×15 pixels. The figure however, demonstrates training only at the last two

levels of resolutions. The test data is a video with 1170 frames which is again transformed to lower resolution data at the mentioned rates.

Table 5.1: Video Statistics

Features	Value
Resolution	640 × 480
Frame Rate	30 per sec
Date Rate	216705kbps
Compression	None

In the second training data set in Figure 5.2 the frames in the test video has 1800 frames and the same set of resolutions as before. The testing sequence is arranged to depict different moments of time where the region of interest changes enabling the foveated pixels to shift their zone of focus just as would be the case if one would observe the behavior of the human eye.

Chapter 6

Concluding Remarks

6.1 Significance of the Result

A method for detecting motion that uses color and structure information is presented. This method extends the Census transform to achieve robust detection by the use of gradient information and a base three encoding technique. Gradient information allows more accurate detection of the outlines of moving objects, while base-three encoding allows to deal effectively with regions of relatively uniform color. The efficiency of the approach is demonstrated by the fact that the entire procedure can be easily carried out in real time.

In this thesis a foveal vision system is also developed that reliably uses motion detection to select foveal areas. Experimental results using a simulated foveal camera show that this selection can be performed reliably. Additionally, the algorithm presented is extremely efficient, consisting mostly of table look-up operations. While the algorithm has relatively high memory requirements due to the need to store two look-up tables for every pixel, the fact that only low-resolution images are processed makes its application using small computers or even embedded hardware feasible.

6.2 Future Work

Future work will include developing a dynamic update of the background model, developing methods for taking advantage of gradient direction information, and exploring other color spaces. One of the other options that can be considered in this case is to implement the Lab color space which can handle illumination changes more effectively than the standard R-G-

B one. Also while the current system performs at two levels of resolution, additional gains in performance can be obtained by having higher processing levels, where, for example, vehicles, pedestrians and faces could be identified and sensed at three different levels of resolution. The plan is to investigate how to best achieve these improvements.

Another aspect that also remains unexplored is the idea of using information sources in addition to color and structure. As a result future work can also include combined information from optical flow methods that can handle faster changes in intensity variations. Again, since such situations are taken into consideration where only important parts of the image are transmitted in high resolution for face or pedestrian detection such combined information can single out regions of interest more effectively. While this could have been an computationally expensive process low resolution imagery allows this to be carried out at an effective rate.

References

- [ACH97] Fabián Arrebola, Pelegrin Camacho, and Francisco Sandoval Hernandez. Generalization of shifted fovea multiresolution geometries applied to object detection. In *International Conference on Image Analysis and Processing*, volume 2, pages 477–484, 1997.
- [BFB04] Csaba Beleznai, Bernhard Fruhstuck, and Horst Bischof. Human detection in groups using a fast mean shift procedure. In *International Conference on Image Processing*, volume 1, pages 349–352, 2004.
- [BS89] Cesar Bandera and P Scott. Foveal machine vision systems. In *IEEE International Conference on Systems, Man and Cybernetics*, volume 2, pages 596–599, 1989.
- [cC98] Ee chien Chang. *Foveation Techniques and Scheduling Issues in Thinwire Visualization*. PhD thesis, New York University, 1998.
- [CD00] R. Cutler and L. S. Davis. Robust real-time periodic motion detection, analysis, and applications. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(8):781–796, 2000.
- [CMY99] E.C. Chang, S. Mallat, and C. Yap. Wavelet foveation. *Journal of Applied and Computational Harmonic Analysis*, 9(3):312–335, 1999.
- [CY97] Ee Chien Chang and Chee K. Yap. A wavelet approach to foveating images. In *Thirteenth annual symposium on Computational geometry*, pages 397–399, 1997.

- [FE04] Bernhard Froba and Andreas Ernst. Face detection with the modified census transform. In *Sixth IEEE International Conference on Automatic Face and Gesture Recognition*, pages 91–96, Erlangen, Germany, May 2004.
- [FF95] Brian V. Funt and Graham D. Finlayson. Color constant color indexing. *IEEE Transaction on Pattern Analysis and Machine Intelligence*, 17(5):522–529, 1995.
- [GP98] Wilson S. Geisler and Jeffrey S. Perry. A real-time foveated multiresolution system for low-bandwidth video communication. In *In Proceedings SPIE*, pages 294–305, 1998.
- [HKR97] B. Heisele, U. Kressel, and W. Ritter. Tracking non-rigid, moving objects based on color cluster flow. In *Proceedings of 1997 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pages 257–260, San Juan, Puerto Rico, June 1997.
- [HN00] Stefan Huwer and Heinrich Niemann. Adaptive change detection for real-time surveillance applications. In *Third IEEE International Workshop on Visual Surveillance*, pages 37–46, Dublin, Ireland, 2000.
- [HSP03] H.Elzein, S.Lakshmanan, and P.Watta. A motion and shape-based pedestrian detection algorithm. In *Intelligent Vehicles Symposium, 2003. Proceedings. IEEE*, pages 500–504, June 2003.
- [JC98] Daesik Jang and Hyung-Il Choi. Moving object tracking using active models. *Proceedings of 1998 International Conference on Image Processing (ICIP 98)*, 3:648–652, Oct 1998.
- [JKM] Barry Brumitt John Krumm, Kentaro Toyama and Brian Meyers. The ultimate futility of background subtraction. *International Journal of Computer Vision*, pages 3–8.

- [JRS06] Agnes Just, Yann Rodriguez, and Marcel Sebastien. Hand posture classification and recognition using the modified census transform. In *Proceedings of the 7th International Conference on Automatic Face and Gesture Recognition (FGR06)*, pages 351–356, 2006.
- [KG96] Philip Kortum and Wilson Geisler. Implementation of a foveated image coding system for image bandwidth reduction. In *In Human Vision and Electronic Imaging, SPIE Proceedings*, volume 2657, pages 350–360, 1996.
- [KGG99] T Kyuel, W Geisler, and J Ghosh. Retinally reconstructed images: digital images having a resolution match with the human eye. *IEEE Transactions on Systems, Man and Cybernetics, Part A: Systems and Humans*, 29(2):235–243, 1999.
- [Kuy98] Turker Kuyel. *Foveated models for compression, texture discrimination and classification*. PhD thesis, University of Texas at Austin, 1998.
- [Lee00] Sanghoon Lee. *Foveated video compression and visual communication over wireless and wireline networks*. PhD thesis, University of Texas at Austin, 2000.
- [LK81] Brian Lucas and Takeo Kanade. An iterative image registration technique with an application to stereo vision. *Proceedings of Imaging of Imaging understanding workshop*, pages 121–130, 1981.
- [LPB98] Sanghoon Lee, M.S. Pattichis, and A.C Bovik. Rate control for foveated mpeg/h.263 video. In *International Conference on Image Processing*, volume 2, pages 365–369, 1998.
- [LPB01] Sanghoon Lee, M.S Pattichis, and A.C Bovik. Foveated video compression with optimal rate control. *IEEE Transactions on Image Processing*, 10(7):977 – 992, 2001.

- [LPB02] Sanghoon Lee, M.S Pattichis, and A.C Bovik. Foveated video quality assessment. *IEEE Transactions on Multimedia*, 4(1):129–132, 2002.
- [MA06] José Martínez and Leopoldo Altamirano. A new foveal cartesian geometry approach used for object tracking. In *Proceedings of the 24th IASTED International Conference on Signal Processing, Pattern Recognition, and Applications*, pages 133–139, Innsbruck, Austria, February 2006.
- [MMPR03] Antoine Monnet, Anurag Mittal, Nikos Paragios, and Visvanathan Ramesh. Background modeling and subtraction of dynamic scenes. In *ICCV '03: Proceedings of the Ninth IEEE International Conference on Computer Vision*, pages 1305–1312, Washington, DC, USA, 2003.
- [MRG99] Stephen J. McKenna, Yogesh Raja, and Shaogang Gong. Tracking colour objects using adaptive mixture models. *Image and Vision Computing*, 17(3/4):225–231, March 1999.
- [NO99] Takayuki Nakamura and Tsukasa Ogasawara. Online visual learning method for color image segmentation and object tracking. *Proceedings of 1999 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS '99)*, 1:222–228, 1999.
- [RC01] Ying Ren and Ching-Seng Chua. Motion detection with non-stationary background. In *Proceedings of the 11th International Conference on Image Analysis and Processing*, Palermo, Italy, September 2001.
- [SB91] M. J. Swain and D. H. Ballard. Color indexing. *International Journal of Computer Vision*, 7(1):11–32, November 1991.
- [SG99] Chris Stauffer and W.E.L Grimson. Adaptive background mixture models for real time tracking. *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2:252, 1999.

- [Sha93] J.M Shapiro. Embedded image coding using zerotrees of wavelet coefficients. *IEEE Transactions on Signal Processing*, 41(12):3445–3462, 1993.
- [Ste04] Fridtjof Stein. Efficient computation of optical flow using the census transform. In *26th Symposium of the German Association for Pattern Recognition (DAGM 2004)*, pages 79–86, Tubingen, August 2004.
- [TS93] Massimo Tistarelli and Giulio Sandini. On the advantages of polar and log-polar mapping for direct estimation of time-to-impact from optical flow. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 15(4):401–410, April 1993.
- [UT02] Akira Utsumi and Nobuji Tetsutani. Human detection using geometrical pixel value structures. *Fifth IEEE International Conference on Automatic Face and Gesture Recognition*, page 39, 2002.
- [VJS03] P. Viola, M. J. Jones, and D. Snow. Detecting pedestrians using patterns of motion and appearance. *IEEE International Conference on Computer Vision*, 2:734–741, 2003.
- [WADP97] Christopher Richard Wren, Ali Azarbayejani, Trevor Darrell, and Alex Pentland. Pfunder: real-time tracking of the human body. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 19(7):780–785, 1997.
- [WOBS94] Richard Wallace, Pin Weng Ong, Ben Bederson, and Eric Schwartz. Space variant image processing. *International Journal of Computer Vision*, 13(1):71–90, 1994.
- [YMAS01] Kunio Yamada, Kenji Mochizuki, Kiyoharu Aizawa, and Takahiro Saito. Motion segmentation with census transform. *Advances in Multimedia Information Processing - PCM 2001*, 2195:903–908, January 2001.

- [ZW94] Ramin Zabih and John Woodfill. Non-parametric local transforms for computing visual correspondence. In *European Conference on Computer Vision*, pages 151–158, May 1994.

Curriculum Vitae

Manali Chakraborty was born on October 29, 1984. She graduated from Carmel High School, Kolkata, India in the summer of 2003 and entered West Bengal University of Technology, India in the fall of 2003. For pursuing her Masters degree she decided to enroll in an U.S. institution. So in the fall of 2007 she started her graduate education in the Computer Science program at the University of Texas at El Paso. While she was a student at UTEP she worked as both a Teaching and Research Assistant. Her work was related to Machine Learning and Computer Vision and she was a part of the Vision Learning and Language Lab at UTEP.

Apart from her research she also took keen interest in sports specially ping-pong and tennis. She was also a member of Computing Research Association for Women and actively participated in activities for women in Computer and Computing Sciences.

Permanent address: 1700 Hawthorne Street, Apt 331

El Paso, Texas 79902