Smart Surveillance System for Cattle Detection

By

Sonali A. Pande

09MCES05

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

AHMEDABAD-382481

May 2012

Smart Surveillance System for Cattle Detection

Major Project

Submitted in partial fulfillment of the requirements

For the degree of

M.Tech. In Computer Science & Enginnering

By

Sonali A. Pande

09MCES05

Guide

Prof. Swati Jain

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

AHMEDABAD-382481

May 2012

UNDERTAKING

I, Sonali A. Pande, Roll No. 09MCES05, give undertaking that the Major Project entitled ”Smart Surveillance System for Cattle Detection” submitted by me, towards the partial fulfillment of the requirements for the degree of Master of Technology in Computer Science and Technology of Nirma University, Ahmedabad, is the original work carried out by me and I give assurance that no attempt of plagiarism has been made. I understand that in the event of any similarity found subsequently with any published work or any dissertation work elsewhere; it will result in severe disciplinary action.

Signature

Date:

Place: Ahmedabad

CERTIFICATE

This is to certify that the Major Project entitled "Smart Surveillance System for Cattle Detection" submitted by Sonali Pande(09MCES05), towards the partial fulfillment of the requirements for the degree of Master of Technology in Computer Science and Engineering of Nirma University of Science and Technology, Ahmedabad is the record of work carried out by her under my supervision and guidance. In my opinion, the submitted work has reached a level required for being accepted for examination. The results embodied in this major project, to the best of my knowledge, haven't been submitted to any other university or institution for award of any degree or diploma.

Prof.SwatiJain
Professor,
Department of Computer Engineering,
Institute of Technology,
Nirma University, Ahmedabad.

Dr.S.N.Pradhan
Professor and PG-Coordinator,
Department of Computer Engineering,
Institute of Technology,
Nirma University, Ahmedabad.

Prof.D.J.Patel
Professor and Head,
Department of Computer Engineering,
Institute of Technology,
Nirma University, Ahmedabad.

Dr.K.Kotecha
Director,
Institute of Technology,
Nirma University, Ahmedabad.

ACKNOWLEDGEMENT

With immense pleasure, I would like to present this report on the dissertation work titled "Smart Surveillance System for Cattle Detection and Identification". It gives me great pleasure in expressing thanks and profound gratitude to **Ms. Swati Jain**, Assistant Professor, Department of Computer Science and Engineering, Institute of Technology, Nirma University, Ahmedabad for her valuable guidance and constant support throughout the project work. I heartily thank her for sparing her precious time to give her suggestions and the clarity of the concepts of the topic that helped me a lot during this study.

I would like to extend my gratitude to **Dr.S.N.Pradhan**, Programme Co-ordinator M.Tech. CSE, Institute of Technology, Nirma University, Ahmedabad whose keen interest and excellent knowledge base helped me to finalize the topic of the dissertation work.

My sincere thanks and gratitude to **Prof. D.J. Patel**, Professor and Head, Computer Engineering Department, Institute of Technology, Nirma University, Ahmedabad for his continual kind words of encouragement and motivation throughout the dissertation work.

I would also like to thanks **Dr. Ketan Kotecha**, Hon'ble Director,Institute of Technology,Nirma University, Ahmedabad for providing basic infrastructure and healthy research environment.

I would like to thank my Institution, all my faculty members in Department of Computer Science and my colleagues without whom this project would have been a distant reality.

And finally I would extend my thanks to my husband **Mr. Ankit Virparia** , for his constant support and help. He encouraged me whenever I was stuck up in my work. I am also thankful to my family members who have helped me in capturing various videos.

- Sonali A. Pande

09MCES05

ABSTRACT

Video object segmentation is an important part of real time surveillance system. For any video segmentation algorithm to be suitable in real time, must require less computational load. The dissertation work presented here is divided into two main parts: (1)moving object detection, (2)identification of cattle from detected moving objects. To detect a moving object the first step required is segmentation. Three different segmentation methods namely K-means clustering algorithm, region growing algorithm and background subtraction method, have been applied on different videos and the results were analyzed. It could be found experimentally that the computation time required for K-means algorithm is very large. Therefore it cannot produce results in real time. Region growing algorithm requires the seed points which indicate the moving object. Hence region growing algorithm is also not applicable in real time. Background subtraction method is fast and produces good results. But it requires a background image from which subsequent frames will be subtracted to obtain the moving object. In order to adapt the background image with the changing video scene, a new algorithm for generating an optimal background image has been proposed. Adaptive background generation is done by evaluating the probability of occurrence of the intensity value at each pixel coordinate. Algorithm has been implemented on videos with various changes in the scene in which the results are quite encouraging. Results obtained with the proposed algorithm are compared with the traditional background subtraction method.

For identification of the cattle feature based method has been used. The algorithm is applied on binary image which is obtained after performing background subtraction. By considering the changes in intensity of the neighboring pixels, a count is maintained which is incremented every time the intensity of the pixel changes from 0 to 255. Keeping a condition on the number of white count obtained can identify whether the moving object under consideration is identified as a cattle or not. The algorithm has been applied on different videos and satisfactory results were obtained.

ABBREVIATION NOTATION

S3 .... Smart Surveillance System
CCTV .... Closed-Circuit Television
VCR .... Video Cassette Recorder
DVR .... Digital Video Recorder
FOV .... Field Of View
FPS .... Frames Per Second
3D .... Three Diamension
FG .... Foreground
BG .... Background
RG .... Region Growing
BGI .... Background Image
THI .... Thresholded Image
RGB .... Red Green Blue
AFD .... Accumulative Frame Difference

Contents

List of Figures

Chapter 1

Introduction

1.1 General Introduction

Video Surveillance systems are widely used these days for a number of applications.Video surveillance, more commonly called CCTV (closed-circuit television), is an industry that is more than 30 years old and one that has had its share of technology changes. The ever increasing demands of the end users, drives the changes in technology. In the video surveillance market, the demands include:[1]

- Better image quality
- Simplified installation and maintenance
- More secure and reliable technology
- Longer retention of recorded video
- Reduction in costs
- Size and scalability
- Remote monitoring capabilities
- Integration with other systems
- More built-in system intelligence

To meet these requirements, video surveillance has experienced a number of technology shifts.Today's systems have come a long way from the early analog tube cameras connected to a VCR. Now-a-days use of network cameras and PC servers for video recording in a fully digitized system has taken its place. However, in between the fully analog and the fully digital systems, there are several solutions which are partly digital. These solutions include a number of digital components but do not represent fully digital systems. In true network video systems, the video is continuously being transported over an IP network, and which are fully scalable and flexible. The advancement in video surveillance can be broadly classified into following three categories:[2]

Fully analog:

- Analog CCTV systems using VCR

Partly digital:

- Analog CCTV systems using DVR
- Analog CCTV systems using network DVR

Fully digital:

- Network video systems using video servers
- Network video systems using network cameras

Analog CCTV systems using VCR

An analog system using a VCR (Video Cassette Recorder) represents a fully analog system consisting of analog cameras with coax output, connected to the VCR for recording. The VCR uses the same type of cassettes as a home VCR. The video is not compressed, and if recording at full frame rate, one tape lasts a maximum of 8 hours. In larger systems, a quad or multiplexer can be connected in between the camera and the VCR. The quad/multiplexer makes it possible to record several cameras to one VCR, but at the cost of a lower frame rate. To monitor the video, an analog monitor is used.

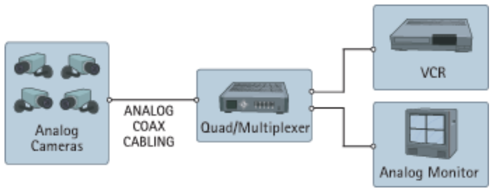

Figure 1.1: Analog CCTV systems using VCR [2]

Analog CCTV systems using DVR

An analog CCTV system using a DVR (Digital Video Recorder) is an analog system with digital recording. In a DVR, the videotape is replaced with hard drives for the video recording, which requires the video to be digitized and compressed in order to store as many day's worth of video as possible. With early DVRs, hard disk space was limited - so recording du-

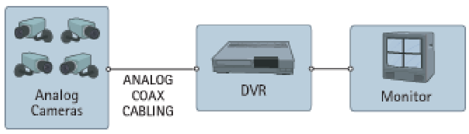

Figure 1.2: Analog CCTV systems using DVR [2]

ration was limited, or a lower frame rate had to be used. Recent development of hard disks means space is no longer a major problem. Most DVRs have several video inputs, typically 4, 9, or 16, which means they also include the functionality of the quad and multiplexers. The DVR system adds the following advantages:

- No need to change tapes
- Consistent image quality

Analog CCTV systems using network DVR

An analog CCTV system using a network DVR is a partly digital system which includes a network DVR equipped with an Ethernet port for network connectivity. Since the video is digitized and compressed in the DVR, it can be transported over a computer network to be monitored on a PC in a remote location. Some systems can monitor both live and recorded

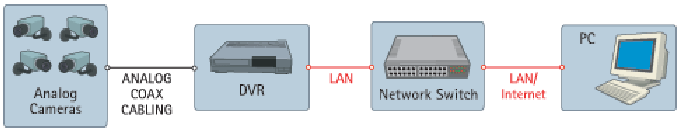

Figure 1.3: Analog CCTV systems using network DVR [2]

video, while some can only monitor recorded. Furthermore, some systems require a special Windows client to monitor the video, while others use a standard web browser; the latter making the remote monitoring more flexible.

The network DVR system adds the following advantages:

- Remote monitoring of video via a PC
- Remote operation of the system

Network video systems using video servers

A network video system using video servers includes a video server, a network switch and a PC with video management software. The analog camera connects to the video server, which digitizes and compresses the video. The video server then connects to a network and transports the video via a network switch to a PC, where it is stored on hard disks. This is a true network video system.

A network video system using video servers adds the following advantages:

- Use of standard network and PC server hardware for video recording and management
- The system is scalable in steps of one camera at a time
- Off-site recording is possible
- It is future-proof since the system can easily be expanded by incorporating network camera

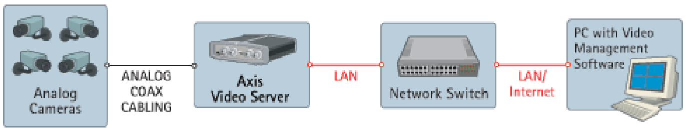

Figure 1.4: Network video systems using video servers [2]

Network video systems using network cameras

A network camera combines a camera and computer in one unit, which includes the digitization and compression of the video, as well as a network connector. The video is transported over an IP-based network, via network switches, and recorded to a standard PC with video management software. This represents a true network video system, and is also a fully digital system, where no analog components are used.
A network video system using network cameras adds the following advantages:

- High resolution cameras (megapixel)
- Consistent image quality
- Power over Ethernet and wireless functionality
- Pan/tilt/zoom, audio, digital inputs and outputs over IP along with video
- Full flexibility and scalability

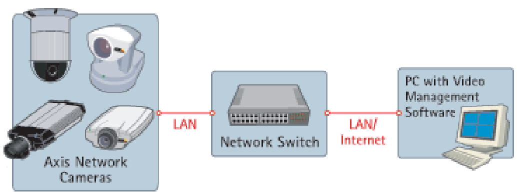

Figure 1.5: Network video systems using network cameras [2]

1.2 Applications of Surveillance System

Following are a few areas where video surveillance systems are used:[3]

Biometric surveillance

Biometric surveillance refers to technologies that measure and analyze human physical and/or behavioral characteristics for authentication, identification, or screening purposes. Examples of physical characteristics include fingerprints, DNA, and facial patterns. Examples of mostly behavioral characteristics include gait (a person's manner of walking) or voice. Facial recognition is the use of the unique configuration of a person's facial features to accurately identify them, usually from surveillance video. Both the Department of Homeland Security and DARPA are heavily funding research into facial recognition systems. Another form of behavioral biometrics, based on affective computing, involves computers recognizing a person's emotional state based on an analysis of their facial expressions, how fast they are talking, the tone and pitch of their voice, their posture, and other behavioral traits. This might be used for instance to see if a person is acting "suspicious" (looking around furtively, "tense" or "angry" facial expressions, waving arms, etc.).

Aerial surveillance

Aerial surveillance is the gathering of surveillance, usually visual imagery or video, from an airborne vehicle-such as a unmanned aerial vehicle, helicopter, or spy plane. Military surveillance aircraft use a range of sensors (e.g. radar) to monitor the battlefield. Digital imaging technology, miniaturized computers, and numerous other technological advances over the past decade have contributed to rapid advances in aerial surveillance hardware such as micro-aerial vehicles, forward-looking infrared, and high-resolution imagery capable of identifying objects at extremely long distances. For instance, the MQ-9 Reaper, a U.S. drone plane used for domestic operations by the Department of Homeland Security, carries cameras that are capable of identifying an object the size of a milk carton from altitudes of 60,000 feet, and has forward-looking infrared devices that can detect the heat from a human body at distances of up to 60 kilometers.

Corporate surveillance

Corporate surveillance is the monitoring of a person or group's behavior by a corporation. The data collected is most often used for marketing purposes or sold to other corporations, but is also regularly shared with government agencies. It can be used as a form of business intelligence, which enables the corporation to better tailor their products and/or services to be desirable by their customers. Or the data can be sold to other corporations, so that they can use it for the aforementioned purpose. Or it can be used for direct marketing purposes, such as the targeted advertisements on Google and Yahoo, where ads are targeted to the user of the search engine by analyzing their search history and emails, which is kept in a database.

1.3 Project Definition

The definition for the project is 'Smart Surveillance System for Cattle Detection'. The main goal of the proposed definition is to develop a prototype consisting of a single camera, for real time detection and identification of cattle form the moving objects. This can be done by mounting a camera at a particular angle and capturing the video. Video is to be processed first for detecting the moving objects and after having detected the moving objects, identifying a cattle out of them.

1.3.1 Description

Video-based surveillance when started with purely analog CCTV system required human operators for the processing of visual information streaming in from often multiple sources. With the shift in technology and also with the massive improvement in these systems, there still remains the complete dependence on human operators. In recent years, demand for real time processing of video has increased. For processing a video in real time requires very fast and robust algorithm. In order to meet this requirement it is necessary that a surveillance system must be smart. The ultimate goal of Smart Surveillance System (S3) is to allow video data to be used for online alarm generation to assist human operators and for offline inspection effectively. Such smart systems are able to generate real-time alarms defined on complex events and handle distributed storage and content-based retrieval of

video data.
The IBM smart surveillance system [5] is one of the few advanced surveillance systems which provides not only automatically monitor a scene but also the capability to manage the surveillance data, perform event based retrieval, receive real time event alerts. The IBM S3 is easily customized to the requirement of different applications. Figure shows the framework of a general smart surveillance system.

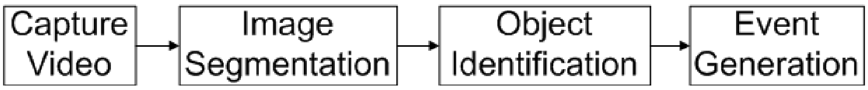

Figure 1.6: Block Diagram of General Smart Video Surveillance System

1.4 Scope of work

Smart Surveillance System for cattle detection and identification aims to detect the moving objects from a video and identifying a cattle from the various objects detected. The work of dissertation will be limited to identification of just cattle and not any other animal or object. It is considered that the camera is stationary. The prototype will produce results for videos which are taken only during day time.

1.4.1 Need for the system

On express highways, the vehicles move at a very high speed. The boundaries of these highways are very low due to which there is a possibility of animals crossing the boundary and coming on the highway. This may result in major accident. In such a scenario the proposed framework would be helpful as it can detect the unwanted object and generate an alarm when it enters the area.
In a huge campus like a college campus, residential area, office building or any other campus this framework will be beneficial.
Another application of this system would be in fields. In villages, farmers are concerned with the destruction of their crop by any cattle. They have to constantly keep a watch to prevent the cattle from entering into the field. If the farmers have computers at their place,

then such system can help then in detecting the presence of a cattle in the field.

1.5 Thesis Organization

The work of Smart Surveillance System for Cattle Detection and Identification is presented in nine chapters.

Chapter two, talks about the existing systems and the literature survey done. It gives a brief description of the related work.

Chapter three, gives an insight to the different methods used for moving object detection.

Chapter four, lists the various challenges faced in implementing the work.

Chapter five, represents a new algorithm for adaptive background image maintenance.

Chapter six, illustrates the algorithm used for performing object identification.

Chapter seven, discusses the results and analysis of different video sequences.

Chapter eight, outlines the conclusion and future scope of this work

Chapter 2

Literature Survey

This chapter covers the studies and work related to the topic. A lot of work has been carried on object detection, tracking and calibration of cameras. A Smart Surveillance System makes use of automatic image understanding technique to extract information from the surveillance data.

Images are considered as one of the most important medium of conveying information. An important aspect of machine learning is to understand the image and to extract information out of it. To understand an image the first step is to segment it and find out different objects in it.Although segmentation is considered to be one of the primary steps in object detection, it is also considered to be one of the most popular problems in computer vision. Byoung [2] in his review of previous related studies, categorized these techniques into following: thresholding approaches, contour based approaches, region based approaches, clustering based approaches and other optimization based approaches using a Bayesian framework, neural networks [2]. A few segmentation methods for object detection have been discussed below.

2.1 Video Surveillance System

Min Moon, Chulho Won, Sung Bum Pan[1] in their paper "Multi-Modal Human Identification based on Smartcard in Video Surveillance System" have suggested a system to protect the privacy of humans using color and height information based on smartcard.

While Surveillance systems can be useful for monitoring inappropriate behaviors, it raises the problem of breaking the privacy of general public. To distinguish human privacy protection in surveillance system, the identification technique is necessary. Human height and clothing color are used for human identification. The system enhances the human identification under the surveillance camera environment using multiple data such as smartcard, clothing-color and height, rather than using single information.

Thiago T. Santos , Carlos H. Morimoto[9] in their paper "Multiple camera people detection and tracking using support integration" have proposed a method to locate and track people by combining evidence from multiple cameras using homography constraint. The proposed method use foreground pixels from simple background subtraction to compute evidence of the location of people on a reference ground plane. The algorithm computes the amount of support that basically corresponds to the "foreground mass" above each pixel. Therefore, pixels that correspond to ground points have more support. The support is normalized to compensate for perspective effects and accumulated on the reference plane for all camera views. The detection of people on the reference plane becomes a search for regions of local maxima in the ccumulator. Many false positives are filtered by checking the visibility consistency of the detected candidates against all camera views. The remaining candidates are tracked using Kalman filters and appearance models. The main contribution of this paper is the definition of a novel algorithm based on the homography constraint that does not rely on single view segmentation of the subjects or previous tracking information.

Chung-Hao Chen a, Yi Yao b, David Page c, Besma Abidi d, Andreas Koschan d, Mongi Abidi[8] in their paper "Camera handoff with adaptive resource management for multi-camera multi-object tracking" have designed a handoff algorithm to obtain a continuously tracked and consistently labeled trajectory of the object of interest in multi-camera surveillance system. Camera handoff is a decision process of transferring a mobile object from one camera to another, wherein consistent labeling solving the identity problem among multiple observing cameras and laying the foundation for camera handoff. In general, camera handoff regulates the collaboration among multiple cameras and answers the questions of When and Who: when a handoff request should be triggered to secure sufficient time for a successful consistent labeling and who is the most qualified camera to take over the object of interest before it falls out of FOV of currently observing camera. Most existing handoff algorithms choose the camera to which the object of interest is approaching. This simple rule is frequently insufficient and leads to unnecessary handoffs.

2.2 Color Image Segmentation

A.V. Wangenheim a, R.F. Bertoldi a, D.D. Abdala b, A. Sobieranski a, L. Coser a, X. Jiang b, M.M. Richter c, L. Priese d, F. Schmitt d[5] in their paper "Color image segmentation using an enhanced Gradient Network Method" have evaluated a new approach for reliable color image segmentation, in images presenting color structure with strong but continuous color changes. Combines an enhanced version of Gradient Network 2, with common region growing method used as pre-segmentation.

Danna Elena and Paul F. Whelen in their paper "Color Image Segmentation Using a Spatial K-Means Clustering Algorithm" have implemented a novel technique to select the dominant colors from the input image using the information from the color histograms. They have generalized the K-means algorithm.

Ali Salem Bin Samma, Rosalina Abdul Salam[4] in their paper "Adaptation of K-Means Algorithm for Image Segmentation" solved the problem of selecting the number of clusters. Added additional steps for convergence step in K-means algorithm.

Bing Leng, Qionghai Dai[15], in their paper "Video Object Segmentation Based on Accumulative Frame Difference" have addressed the problem of extracting video objects from head-shoulder video sequences. They have proposed a method based on Accumulative Frame Difference (AFD). Initially by performing motion analysis on each block of frame, the blocks with fast moving edges are detected. Then, for each block, they have accumulated the frame difference with a different amount of frames, based on its motion attributes. After thresholding and post processing, the objects are obtained.

Yasira Beevi C P, Dr. S. Natarajan[16], in their paper "An efficient Video Segmentation Algorithm with Real time Adaptive Threshold Technique" have presented a video segmentation algorithm for MPEG-4 camera system with change detection, background registration techniques and real time adaptive threshold techniques. Their work also consist of a shadow cancellation mode which can deal with light changing effects and shadow effects.

2.3 Object Identification

Object identification is one of the most fascinating abilities that humans easily possess since childhood. With a simple glance of an object, humans are able to tell its identity or category despite of the appearance variation due to change in pose, illumination, texture, deformation, and under occlusion. Furthermore, humans can easily generalize from observing a set of objects to recognizing objects that have never been seen before. For example, kids are able to generalize the concept of "chair" or "cup" after seeing just a few examples. Nevertheless, it is a daunting task to develop vision systems that match the cognitive capabilities of human beings, or systems that are able to tell the specific identity of an object being observed. The main reasons can be attributed to the following factors: relative pose of an object to a camera, lighting variation, and difficulty in generalizing across objects from a set of exemplar images. Central to object recognition systems are how the regularities of images, taken under different lighting and pose conditions, are extracted and recognized. In other words, all these algorithms adopt certain representations or models to capture these characteristics, thereby facilitating procedures to tell their identities. Few approaches for object recognition are:

2.3.1 Geometry-based approaches

Early attempts on object recognition were focused on using geometric models of objects to account for their appearance variation due to viewpoint and illumination change. The main idea is that the geometric description of a 3D object allows the projected shape to be accurately predicted in a 2D image under projective projection, thereby facilitating recognition process using edge or boundary information (which is invariant to certain illumination change). Much attention was made to extract geometric primitives (e.g., lines, circles, etc.)

that are invariant to viewpoint change [12]. Nevertheless, it has been shown that such primitives can only be reliably extracted under limited conditions (controlled variation in lighting and viewpoint with certain occlusion). An excellent review on geometry-based object recognition research by Mundy can also be found in [11].

2.3.2 Appearance-based algorithms

In contrast to early efforts on geometry-based object recognition works, most recent efforts have been centered on appearance-based techniques as advanced feature descriptors and pattern recognition algorithms are developed [13]. Most notably, the eigenface methods have attracted much attention as it is one of the first face recognition systems that are computationally efficient and relatively accurate [14]. The underlying idea of this approach is to compute eigenvectors from a set of vectors where each one represents one face image as a raster scan vector of gray-scale pixel values. Each eigenvector, dubbed as an eigenface, captures certain variance among all the vectors and a small set of eigenvectors captures almost all the appearance variation of face images in the training set. Given a test image represented as a vector of gray-scale pixel values, its identity is determined by finding the nearest neighbor of this vector after being projected onto a subspace spanned by a set of eigenvectors. In other words, each face image can be represented by a linear combination of eigenfaces with minimum error, and this linear combination constitutes a compact reorientation.

2.3.3 Feature-based algorithms

The central idea of feature-based object recognition algorithms lies in finding interest points, often occurred at intensity discontinuity, that are invariant to change due to scale, illumination and affine transformation. The scale-invariant feature transform (SIFT) descriptor, proposed by Lowe, is arguably one of the most widely used feature representation schemes for vision applications[13]. The SIFT approach uses extrema in scale space for automatic scale selection with a pyramid of difference of Gaussian lters, and keypoints with low contrast or poorly localized on an edge are removed. Next, a consistent orientation is assigned to each keypoint and its magnitude is computed based on the local image gradient histogram, thereby achieving invariance to image rotation. At each keypoint descriptor, the contribution of local image gradients are sampled and weighted by a Gaussian, and then represented by orientation histograms. For example, the 16x16 sample image region and 4x4 array of histograms with 8 orientation bins are often used, thereby providing a 128-dimensional feature vector for each key point. Objects can be indexed and recognized using the histograms of key points in images.

Chapter 3

Challenges

This chapter covers major challenges encountered for the implementation of the proposed definition. The main goal of our definition is to detect and identify a cattle in a video sequence. The segmentation algorithms were initially applied on a single image and then it is applied on a video sequence.

Various challenges are listed as follows:

a. First and most important is detection of the moving object. Segmentation of the object is most important but also has many problems. We first applied k-means algorithm on the image. The challenge here is to decide the number of clusters and the initial center points.

b. Background subtraction method requires that the background image must adapt with the changes in the background scene. Dealing with adaptive background image is another challenge.

c. Identification of cattle out of the various moving objects detected is also a challenge.

Chapter 4

Moving Object Detection

For any video surveillance system the first and foremost task is to detect moving objects. Therefore the first step of processing the video is segmentation. Segmenting an image will result in the image being divided into various regions. It separates the foreground objects from the background. When an appropriate segmentation algorithm is chosen, it will generate desired results. Results generated by segmentation are used for further processing. Therefore the segmentation algorithm must be chosen properly and in accordance with the need of the system for which results have to be generated. A few of the techniques for image segmentation are as follows:

(1) Thresholding

(2) Region growing

(3) Classifiers

(4) Clustering

(5) Markov random fields model

(6) Deformable models

As a part of our dissertation, we have implemented three techniques for image segmentation.

4.1 K-means Clustering Algorithm

The clustering approaches can be categorized into two general groups: partitional and hierarchical clustering algorithms. Partitional clustering algorithms such as K-means and EM clustering are widely used in many applications such as data mining, compression, image segmentation, and machine learning. Therefore, the advantage of clustering algorithms is that the classification is simple and easy to implement. Similarly, the drawbacks are of how to determine the number of clusters and decrease the numbers of iteration.
The K-Means is a nonhierarchical clustering technique that follows a simple procedure to classify a given data set through a certain number of K clusters that are known a priori. The K-Means algorithm updates the space partition of the input data iteratively, where the elements of the data are exchanged between clusters based on a predefined metric (typically the Euclidian distance between the cluster centers and the vector under analysis) in order to satisfy the criteria of minimizing the variation within each cluster and maximizing the variation between the resulting K clusters.
Steps of the classical K-Means clustering algorithm:

(1) Initialization - generate the starting condition by defining the number of clusters and randomly select the initial cluster centers.

(2) Generate a new partition by assigning each data point to the nearest cluster center.

(3) Recalculate the centers for clusters receiving new data points and for clusters losing data points.

(4) Repeat the steps 2 and 3 until a distance convergence criterion is met.

As mentioned before, the aim of the K-Means is the minimization of an objective function that samples the closeness between the data points and the cluster centers, and is calculated as follows:

$$J = \sum_{j=1}^{k} \sum_{i=1}^{n} \left| x_i^{(j)} - c_j \right|^2 \tag{4.1}$$

where $\left| x_i^{(j)} - c_j \right|^2$ is the distance (usually the Euclidian metric) between the data point and the cluster center . It is sensitive to the initialization process that selects the initial cluster

Figure 4.1: Colored images and their corresponding grayscale images

centers (usually randomly picked from input data).

Figure 4.1 shows the frames from a video on which the k-means clustering algorithm is

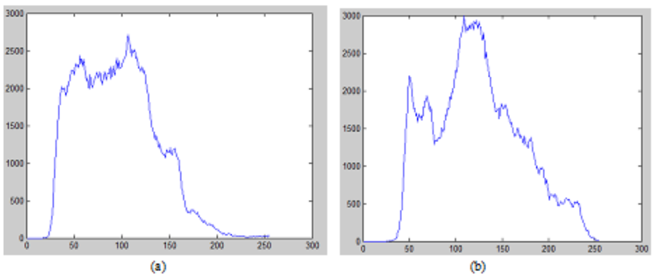

Figure 4.2: (a) Histogram of first image, (b) Histogram of second image

implemented. Figures following it, shows the obtained results. The algorithm has been tested on two different video sequences and also by considering two different cluster sizes. Figure 4.3(a) shows the result of segmentation considering cluster size as 5 and Figure 4.4(b) shows the result considering the cluster size as 4.

From the above results it can be identified that K-means clustering algorithm does not produce appropriate results when applied on complex textured image. Segmentation of complex color-textured images is restricted by two problems. The first problem is generated

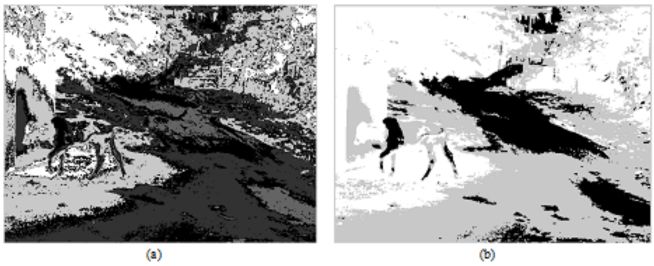

Figure 4.3: Result of segmentation using k-means (a) cluster size=5 (a) cluster size=4 Time=60.035sec

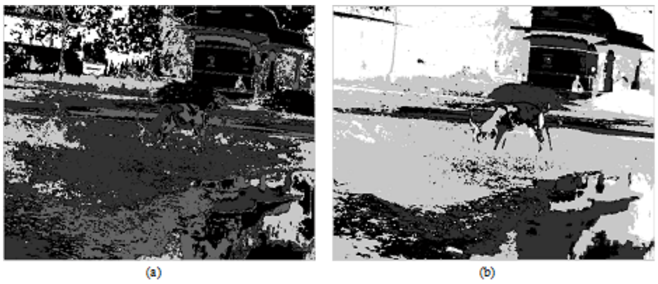

Figure 4.4: Result of segmentation using k-means (a) cluster size=5 (b) cluster size=4 Time=62.39sec

by the starting condition (the initialization of the initial cluster centers), while the second is generated by the fact that no spatial (regional) cohesion is applied during the space partitioning process.

4.2 Region Growing Algorithm

The first region growing method was the seeded region growing method. This method takes a set of seeds as input along with the image. The seeds mark each of the objects to be segmented. The regions are iteratively grown by comparing all unallocated neighboring pixels to the regions. The difference between a pixel's intensity value and the region's mean,δ, is used as a measure of similarity. The pixel with the smallest difference measured this way is allocated to the respective region. This process continues until all pixels are allocated to a region. Seeded region growing requires seeds as additional input. The segmentation

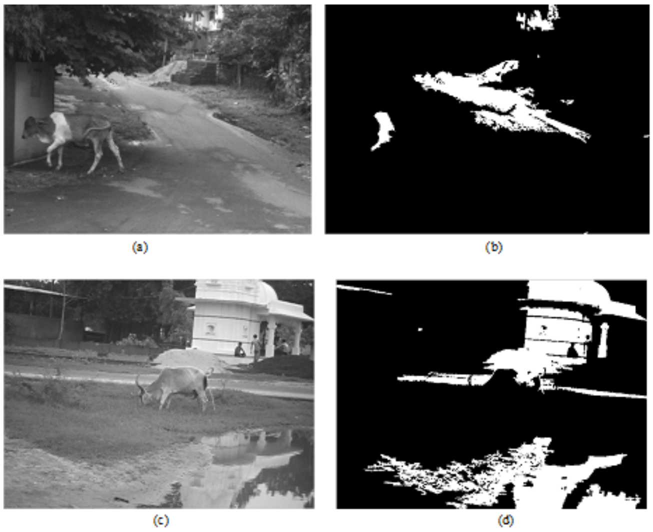

Figure 4.5: video frames and their results after applying region growing algorithm

results are dependent on the choice of seeds. Noise in the image can cause the seeds to be poorly placed. Unseeded region growing is a modified algorithm that doesn't require explicit seeds. Figure 4.4 shows the results obtained on applying the region growing algorithm. The results depict that the region growing algorithm does not produce appropriate results in real time. Region-growing segmentation methods present two main features that limit their applicability for dealing efficiently with natural scenes:[5]

- A static region similarity concept: where pixels or textures within a region are expected to be homogeneous. Typical natural scenes, however, show strong continuous variations of color, presenting a different, dynamic order that is not taken into account by such algorithms.

- Increase in complexity to present more stable results: which usually demands complex computations to detect segment-correlation clues, or are built upon additional texture information. This slows down considerably the processing time without being much more stable when extreme color variations are present.

4.3 Background Subtraction Method

A simple and common motion detection method involves subtracting each new image $I_t(x, y)$ in a video sequence from a model of the background scene void of objects $B_t(x, y)$ and thresholding the resulting difference, highlighting the foreground pixels[3].

The pixels having the values above threshold are classified as foreground pixels. As a result

Figure 4.6: (a)Background image, (b)frame of a video

a binary image is formed.

$$|I_t\,(x, y) - B_t\,(x, y)| > T \tag{4.2}$$

where T is a predefined threshold. It is necessary to update the background image frequently in order to guarantee reliable motion detection.

Figure 4.7: Result of background subtraction

Figure4.7 shows the resultant image obtained on applying background subtraction method. It is required to update the background every hour in order to get appropriate results. Figure4.9 shows six different frames of a video sequence.

Figure 4.8: Background frame of the video sequence

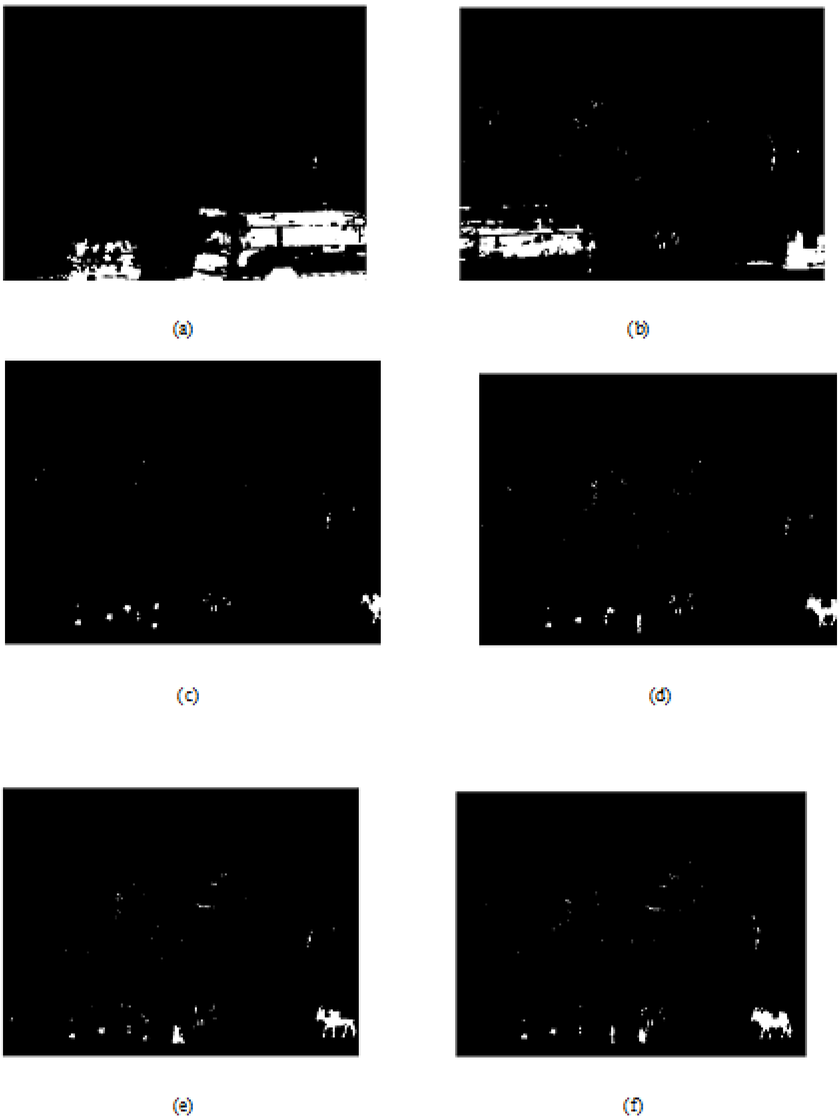

Figure 4.9: Frames containing some moving object

Chapter 5

Generating an Adaptive Background Image

A background image can be considered as an image containing objects which remain inactive in the frame. In an indoor environment, it may consist of the furniture, television, telephone etc. which are non-living things and are not capable of changing their position on their own. In an outdoor environment, it may consist of buildings, electric poles and hoardings which are also non-living objects. These can be considered as static object in the background image. But the appearance of background image may change with time. Following can be the situation where background can change:

(1) An object remains stationary for a long duration of time.

(2) An object initially moving but then becomes stationary for a long time.

(3) An object which was stationary, starts moving after sometime.

In both of these situations, the background image changes. If it is not updated with the changes in the scene then object might not be detected appropriately. To handle both the above mentioned situations a new approach for generating the background image has been proposed. A modified background subtraction method is introduced which has the ability to generate a new reference image with the changing scene. The proposed algorithm runs in parallel with the actual background subtraction method without interrupting the process of moving object detection. The algorithm can generate an image consisting of all the static objects in the scene. At regular interval the algorithm can be repeated to generate a new image and this newly generated image can be used as the background image for carrying out background subtraction method to detect moving objects. When a new background image is generated, it is made available to the object detection process (as shown in Figure5.1) and further subtraction is carried out considering the newly generated image as the background image. As a result the execution time is not affected. The proposed algorithm is able to handle the situations mentioned which cannot be handled by the traditional background subtraction method where the reference background image does not adapt with the changes is the background scene. Implementation of the algorithm considers two cases: (1). image with static background, (2). image with dynamic background. The results which are obtained are compared with the results of applying traditional background subtraction method. Figure 5.1 shows the block diagram of proposed algorithm for generating background image dynamically from a video sequence. Algorithm of the method is as below:

(1) Read the first N frames from the video.

(2) Store all the intensity values of the respective coordinates from all the N frames.

(3) Calculate the probability distribution PD for each coordinate point showing the probability of occurrence of the intensity values at that coordinate.

(4) From the PD find out the intensity value with the maximum probability.

(5) Generate a new background image by placing the maximum intensity values at the respective coordinate point.

(6) Use the newly generated background image for performing image subtraction.

The above mentioned algorithm can be repeated at regular interval in order to obtain a new reference image. The algorithm has to be repeated periodically in order to change the background image which is used for image subtraction when an object which was stationary in the scene has now moved out of the scene.

Implementation of the modified background subtraction method can be subdivided into three main parts: background image generation, background subtraction, background updation.

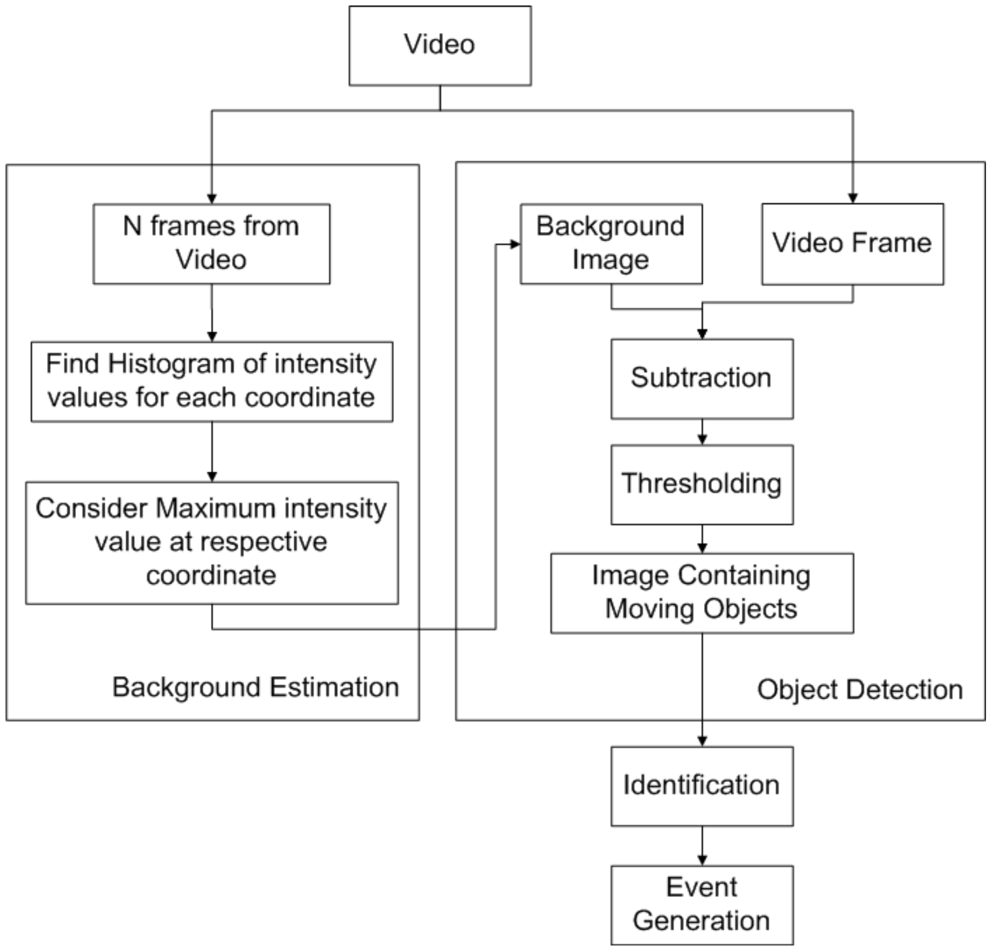

Figure 5.1: Block diagram of the proposed algorithm for generating the background image.

5.1 Background Image Generation

The first and important step for performing background subtraction is to obtain an appropriate background image which does not contain any moving objects.

Consider that a video sequence has N frames. Out of that first T frames are considered for generating a background image. Let $F_t(i, j)$ be a frames from a video where, t=1,2,3,,T and i and j represents the resolution of the frame. The following equation is used to generate a background image:

$$PD(i, j, k) = \frac{1}{t} \sum_{t=1}^{N} F_t(i, j) \tag{5.1}$$

where k is the intensity value.

$$BGI(i, j) = max[PD(i, j, k)] \tag{5.2}$$

BGI(i,j) is the generated background image.

5.2 Background Subtraction

Having got the background image, background subtraction can be applied to detect the moving objects from different video frames. Let N be the total number of frames in the video. BGI(i,j) is the background image. Each new frame of the video is subtracted from BGI(i,j).

$$FD_n(i, j) = |BGI(i, j) - I_n(i, j)| \tag{5.3}$$

where n=1,2,3,,N and $FD_n(i, j)$ is the frame difference of each frame from the background image.

$$THI_n(i, j) = 1, ifFD \geq Th0, ifFD \leq Th \tag{5.4}$$

where Th is the threshold value and $THI_n(i, j)$ is the thresholded image.

5.3 Background Updation

Updating the background image is required to adapt the reference image with the changes in the background scene. To detect a moving object accurately it is required to again generate a new background image. For that the procedure given in Equations (5.1),(5.2),(5.3) are again repeated. This is done parallel to the background subtraction process.

Let BGI(x,y) is the newly generated background image. Now this image can be used for further processing.

Chapter 6

Identifying Cattle Out of Moving Objects

Object detection process will detect all the objects which are moving. But just detection of the moving object is not sufficient. Out of different moving objects it is required to identify the cattle. Object recognition - in computer vision, is the task of finding a given object in an image or video sequence. Humans recognize a multitude of objects in images with little effort, despite the fact that the image of the objects may vary somewhat in different viewpoints, in many different sizes / scale or even when they are translated or rotated. Objects can even be recognized when they are partially obstructed from view. Object recognition in humans is largely invariant with regard to changes in the size, position, and viewpoint of the object. To develop a system which matches the capabilities of human is a challenge. Such a system must consist of the following capabilities:[4]

- Rotational Invariant
- Size Invariant
- Translation Invariant

An identification system must be capable of handling the above three features.
Object recognition methods has the following applications:[5]

- Image panoramas

- Image watermarking
- Global robot localization
- Face detection
- Optical Character Recognition
- Manufacturing Quality Control
- Content-Based Image Indexing
- Object Counting and Monitoring
- Automated vehicle parking systems
- Visual Positioning and tracking
- Video Stabilization

As a part of our dissertation work we have used Feature-based method to identify cattle from different moving objects.

6.1 Feature-based Method

In feature based method for object identification, a search is used to find feasible matches between object features and image features. The primary constraint is that a single position of the object must account for all of the feasible matches. We consider four legs of a cattle as one feature in the lower half of the cattle and its hump and horns as second feature in the upper half of the body. Figure 7.1 shows the block diagram of the algorithm.
Algorithm for cattle identification based on its features:

(1) Find the boundary coordinates of all the moving objects detected in a binary image.

(2) Perform raster scan of the object to find out intensity discontinuity.

(3) Keep a count of the number of times intensity value changes.

(4) If the value of count satisfies the condition for feature matching then the object is identified as a cattle.

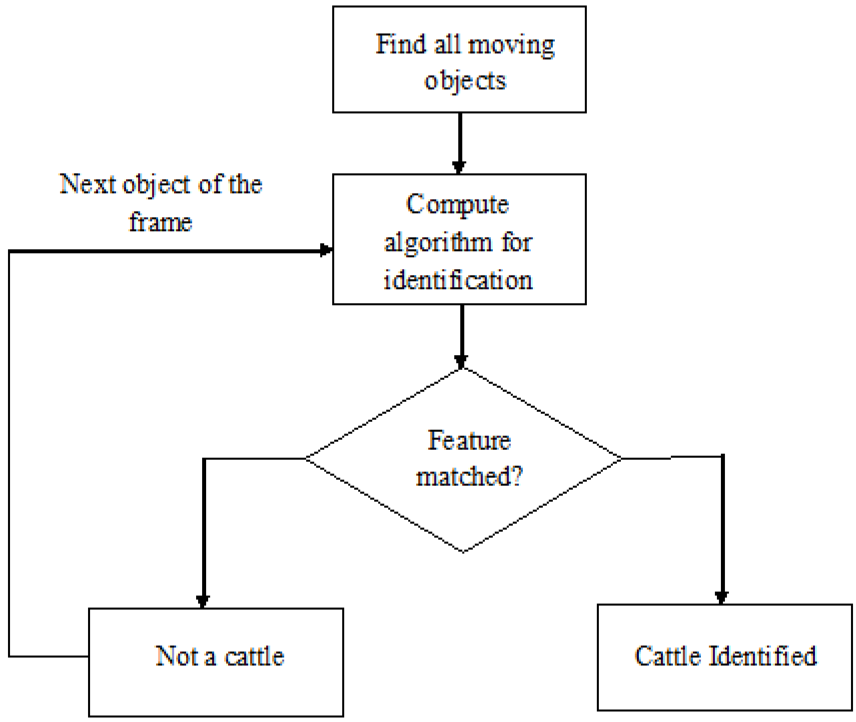

Figure 6.1: Block diagram for cattle identification algorithm

(5) Repeat steps 2,3 and 4 for all the objects detected in a frame.

Chapter 7

Experimental Setup and Results

In the work presented here, Matlab 10 has been used for implementing the algorithms. It is considered that the camera is stationary. Videos have been recorded using a single camera with the resolution of 320x240.

7.1 Moving object detection

7.1.1 Using proposed algorithm

Figure 7.1: Frames from a video (a) a car is stationary, (b) car started moving, (c) car moving farther, (d) car moved out of the frame.

Three different scenarios have been considered on which the algorithm has been applied.

Figure 7.1 shows the frames of a video in which a car is stationary for a while and then it moves out of the frame. The algorithm generates a background image by analyzing the initial few frames of the video. Figure 7.2 shows the generated background image. The

Figure 7.2: Generated background image using proposed algorithm.

background image thus generated is used as the reference image for performing background subtraction. Figure 7.3 shows the frame having moving object and the image obtained after doing subtraction and thresholding.

Figure 7.3: Result of background subtraction (a) video frame in which car has moved out, (b) thresholded image

As the background image does not contain the temporary static object, moving object can be detected accurately. Figure 7.4 shows the scenario in which a car remains stationary for a long duration of time.

The background image generated from the video must contain the stationary object as it does not move for a long duration of time. Figure 7.5 shows the generated background

(a) (b)

Figure 7.4: Frames from a video in which car is stationary.

image.

Figure 7.5: Generated background image using proposed algorithm.

Figure 7.6 shows the result of background subtraction using the background image generated by the algorithm.

7.1.2 Using traditional background subtraction method

Applying traditional background subtraction technique requires having a background image which is used during subtraction. Figure 7.7 shows the first frame of the video which is considered as background image. The background image here contains a stationary object as it was stationary for along duration of time.

Figure 7.8 shows changes in the resultant image obtained after background subtraction as the scene changes.

Figure 7.6: Result of background subtraction (a) video frame containing stationary and moving object both, (b) thresholded image

Figure 7.7: Background image for traditional background subtraction

Analysis: From the results obtained on applying the proposed algorithm it can be analyzed that the algorithm detects moving objects more accurately. The stationary object moves away gradually from the video frames shown in Figure 7.1 therefore the background image generated from the initial few frames does not contain the stationary object. Thus the thresholded image shown in Figure 7.3 detects the moving object correctly. As the car in Figure 7.6(a) remains stationary and the background image generated also contains the stationary object, resultant image obtained after background subtraction detects only the moving object. In traditional background subtraction method, the background image does not change therefore it may not always produce proper results. Figure 6.8(b) detects the moving objects appropriately. But Figure 6.8(d),(f) detects the moving object incorrectly. In Figure 6.8(c),(e) there is no stationary object but still its thresholded image detects a moving object as this frame is subtracted from the background image containing the stationary object. In such cases if the background image does not adapt to the changes in

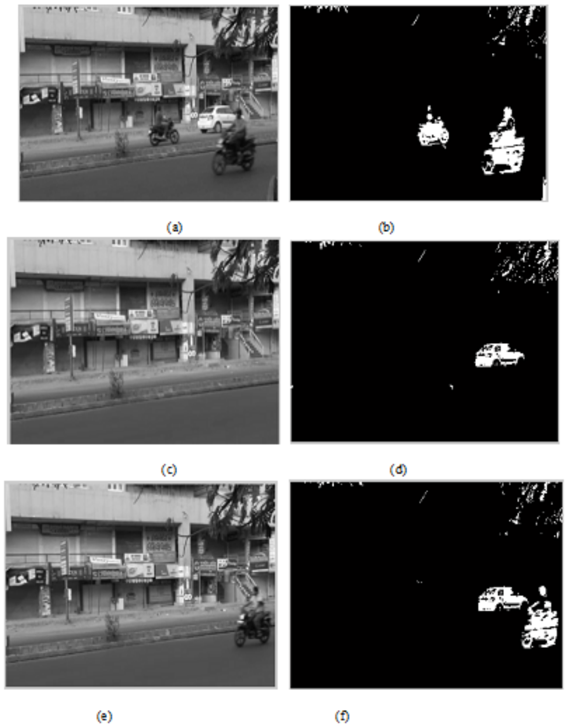

Figure 7.8: Result of background subtraction (a) video frame containing stationary and moving object both, (b) its thresholded image, (c) frame which does not have stationary object, (d) its thresholded image, (e) frame containing only the moving object, (f) its thresholded image.

the scene it will not detect moving objects correctly.

7.2 Object Identification

After detecting the moving objects, identify a cattle from it. The proposed algorithm for object identification has been applied on two different videos.

7.2.1 Results for Video-1

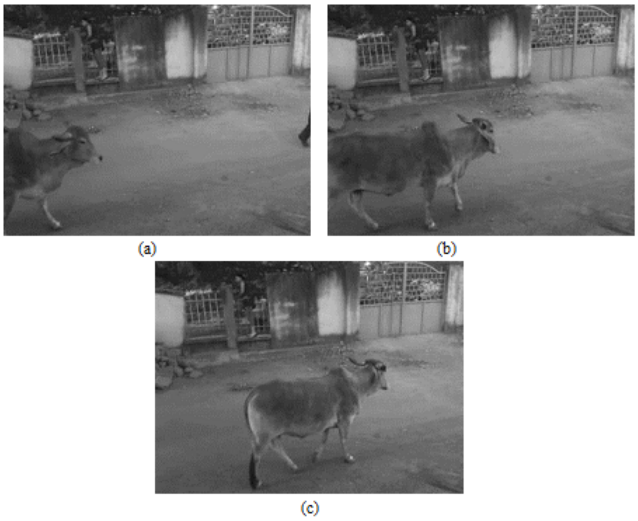

Figure 7.9: different frames from Video-1

The background images generated from the video on applying the proposed algorithm is shown in Figure 7.10

Figure 7.10: Generated background image

Figure 7.11 shows the image obtained after background subtraction and also the image obtained after applying algorithm for cattle identification.

Figure 7.11(c) shows the moving object detected in the frame after applying background subtraction. Figure 7.11(d) shows the plot of white count which is incremented in the

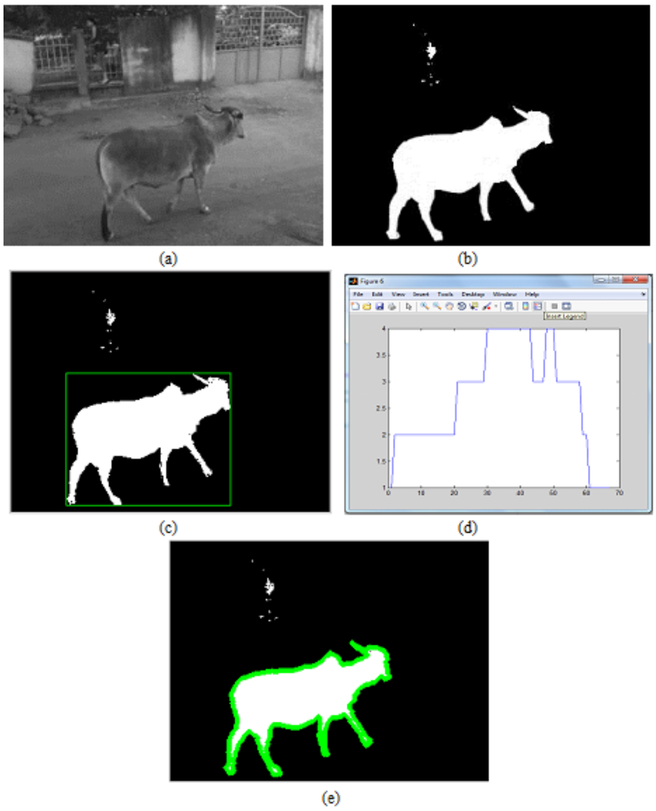

Figure 7.11: Result obtained after applying the algorithms on Video-1

algorithm whenever there is a change in pixel intensity. Figure 7.11(e) shows the final image which shows the identified object.

7.2.2 Results for Video-2

The background images generated from the video on applying the proposed algorithm is shown in Figure 7.13

Figure 7.14 (a),(b) shows the frame and its background subtracted image. Figure 7.14(c) shows the detected moving object. As the moving algorithm is not a cattle, the matching algorithm will not identify it as a cattle.

Figure 7.14 shows the resultant image obtained after background subtraction.

Figure 7.12: Three different frames from Video-2

Figure 7.13: Generated background image

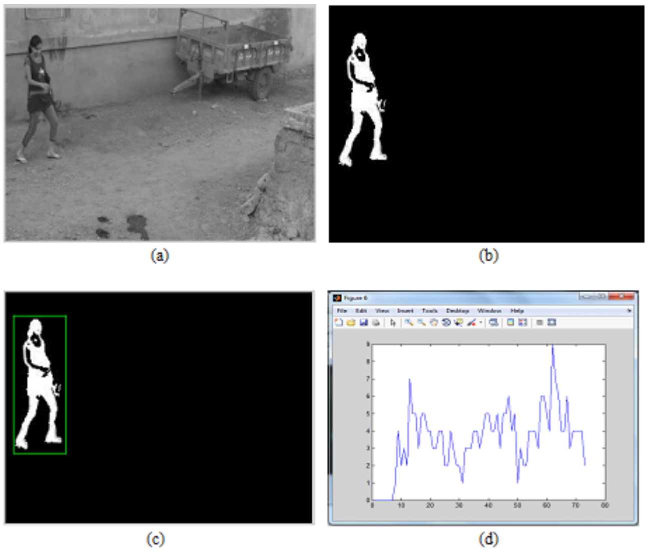

Figure 7.14: obtained after applying the algorithms on Video-2

Chapter 8

Conclusion and Future Work

8.1 Conclusion

Moving object detection and identification involves two important parts: object detection and object identification. Detection of object is done by segmentation. Background subtraction method is used for detection of moving objects. As object identification depends on the result of object detection, it is necessary to have optimal background subtraction method with adaptive background image maintenance. The proposed algorithm for evaluating adaptive background image gives good results when tested with videos in the following scenarios: videos containing only static background, videos containing dynamic background. Real time video applications require fast and robust processing. The proposed algorithm can run parallel can generate the background image without interfering with the object detection process. Using the proposed algorithm moving objects can be detected more accurately. For cattle identification feature based method is used. Features of cattle like four legs, hump and horns are identified from the object. Satisfactory results have been obtained on applying both the algorithms on different video sequences.

8.2 Future Work

(1) Tracking of the cattle to determine the direction of its motion to raise the alarm to appropriate station.

(2) A module can be developed and integrated, which would allow to add the object required to be identified.

Web References

[1]. http://www.axis.com/products/video/about_networkvideo/evolution.htm

[2]. http://www.infosectoday.com/Articles/Video_Surveillance_Systems.htm

[3]. http://en.wikipedia.org/wiki/Surveillance

[4]. http://www.pigeon.psy.tufts.edu/avc/kirkpatrick/default.htm

[5]. http://en.wikipedia.org/wiki/Outline_of_object_recognition

References

[1] Hae-Min Moon, Chulho Won, Sung Bum Pan, "The multi model human identification based on smartcard in video surveillance", 2010 IEEE/ACM International Conference on Green Computing and Communication

[2] B. Jeon, Y. Yung and K.Hong "Image segmentation by unsupervised sparse clustering," pattern recognition letters 27science direct,(2006) 1650-1664

[3] A. Elgammal, R.Duraiswami, D. Harwood, and L.S. davis, "Background and foreground modeling using nonparametric kernel density estimation for visual surveillance", Proceedings of IEEE, 90(7):1151-1163, July 2002

[4] Ali Salem Bin Samma, Rosalina Abdul Salam, "Adaptation of K-Means Algorithm for Image Segmentation", World Academy of Science, Engineering and Technology 50 2009

[5] A.V. Wangenheim a, R.F. Bertoldi a, D.D. Abdala b, A. Sobieranski a, L. Coser a, X. Jiang b, M.M. Richter c, L. Priese d, F. Schmitt d, "Color image segmentation using an enhanced Gradient Network Method", Pattern Recognition Letters 30 (2009) 1404-1412

[6] Jeferson R. Silva, Thiago T. Santos, Carlos H. Morimoto, "Automatic camera control in virtual environments augmented using multiple sparse videos", Computers Graphics 35 (2011) 412-421

[7] Y. J. ZHANG, "A SURVEY ON EVALUATION METHODS FOR IMAGE SEGMENTATION*", Pattern Recoynition, Vol. 29, No. 8, pp. 1335-1346, 1996

[8] Chung-Hao Chen a, Yi Yao b, David Page c, Besma Abidi d, Andreas Koschan d, Mongi Abidi , "Camera handoff with adaptive resource management for multi-camera multi-object tracking", Image and Vision Computing 28 (2010) 851-864

[9] Thiago T. Santos, Carlos H. Morimoto, " Multiple camera people detection and tracking using support integration", Pattern Recognition Letters 32 (2011) 47-55

[10] Konstantinos Tzevanidis a,b, Antonis Argyros, "Unsupervised learning of background modeling parameters in multicamera systems", Computer Vision and Image Understanding 115 (2011) 105-116

[11] J. Mundy. Object recognition in the geometric era: a retrospective. In J. Ponce, M. Hebert, C. Schmid, and A. Zisserman, editors,Toward category-level object recognition, pages 329. Springer-Verlag, 2006.

[12] J. Mundy and A. Zisserman, editors. Geometric invariance in computer vision. MIT Press, 1992.

[13] D. Lowe. Distinctive image features from scale-invariant keypoints. International Journal of Computer Vision, 60(2):91110, 2004.

[14] M. Turk and A. Pentland. Eigenfaces for recognition. Journal of Cognitive Neuroscience, 3(1):71,86, 1991.

[15] Bing Leng , Qionghai Dai, "Video Object Segmentation Based on Accumulative Frame Difference", unpublished.

[16] Yasira Beevi C P, Dr. S. Natarajan, "An efficient Video Segmentation Algorithm with Real time Adaptive Threshold Technique", International Journal of Signal Processing, Image Processing an d Pattern Recognition Vol. 2, No.4, December 2009.

Index