ISSN:0975-9646

Snehal Anwekar et al, / (IJCSIT) International Journal of Computer Science and Information Technologies, Vol. 7 (2) , 2016, 655-660

Flexible Content Based Video Surveillance System for crime Prevention based on moving object detection

Snehal Anwekar[#1], Isha Walimbe[2], Priyanka Suryagan[#3],Alisha Gujarathi[#4], Kalpana Thakre[#5]

[#12345]*Department of Information Technology,*
Savitribai Phule Pune University
Pune, Maharashrta, India

***Abstract*— In recent years, people are more concerned about security. Therefore demand for surveillance systems and the need to process video frames has increased. The description of video via traditional text recording or annotation is not suitable because it cannot fully describe all the contents of surveillance video. Recording data and then processing it can take time. In this paper we propose a system which will be designed to aid crime detection and prevention by processing video frames in real time. Video retrieval is the process of retrieving important information from videos in order to classify them and use them for various purposes. Video Retrieval is an evolved branch of Image retrieval, which works on a broader scale by processing multiple images (frames) at a time. The classification is done according to features which are an integral part of video frames. Features can include color, shape, edges, texture etc.**

This paper proposes a system which is used for detecting motion from a static camera.. It also illuminates certain other similar system surveys. All these systems are divided into two major categories, Active systems and passive systems. The paper concentrates on extracting information from video frames and using it for crime prevention purposes. The system will take real time video and find out anomalies by detecting motion. The system will raise an alarm to notify authorities of unwanted presence.

***Keywords*— Threshold, video surveillance, motion detection, pre-processing, motion detection, retrieval, blob detection.**

I. Introduction

Nowadays, solutions for content-based retrieval of surveillance data are ever more required as the number of video surveillance systems and the amount of stored data drastically increase. The main reason for searching video surveillance data are forensic i.e. to look for video evidence after an incident occurred and infrastructure maintenance [1] .Video surveillance has been a key component in ensuring security at airports, banks, and houses. The proposed system captures live feed and detects the undesired motion. The system is not based on post content based processing and indexing of archived footage, but runs real time filters on live footage and signals unusual events [2]. This reduces the need to actively and constantly watch the live feeds [2]. This system focuses mainly on home surveillance and provides security by raising alarms via text message.

The live feed is processed by applying grayscaling and blurring techniques to remove noise within the image which further undergoes thresholding and blob detection stages after which the motion is detected. The task of extracting moving objects from a video is a fundamental problem of this system. The surveillance video cameras are indeed often fixed, so that the sequences present a static background together with mobile objects that are semantically meaningful [1]. Moving object detection is the fundamental step and it handles segmentation of moving objects from stationary background objects. Commonly used techniques for object detection are background subtraction, Statistical differencing and temporal models. Background subtraction used in this system is done by comparing each frame to a model of static background. After detection of motion, alarm is raised to notify the owner of the system about the crime.

The rest of this paper is organized as follows: Section 2 describes the related work and surveys, Section 3 describes the Proposed System, and Section 4 gives the UML diagrams and flow of system.

II. Related Work

Jerome Meessen et al. [1] proposed a dissimilarity measure for surveillance videos. It is the method in which the similarity between each key-frame Fi and user's query Q is calculated. A measure of their relative dissimilarity, or distance, D (Fi, Q), is proposed. For long video sequences a key-frame is selected to reduce the amount of data to be processed. For each period when the number of objects is constant, the middle frame is selected as its key-frame. The main drawback of this system is that it is static and cannot be adapted to user's needs. In this case another method is used in which distance between each frame and last selected key frame is sequentially measured. If the measured distance exceeds the decided threshold then it is considered as end of a homogeneous scene is reached and a new key frame is selected.

Levente Kovacs1 et al. [2] proposed a video surveillance and event detection and annotation framework for semi-supervised surveillance use. Current detector filters, and the extendable modular interface is used to detect motion. Current filters include local and global unusual motion detectors, left object detector, motion detector, tampering/failure detector, etc. All filters run automatically and do not need any manual handling. This system consists of functions which are all automatic and consist of a panorama image creator from panning camera feeds, mask able motion detector, camera jump detector for cameras that iterate among different stationary positions, unusual global and local motion detector, fight detector, left object detector, camera fail/tampering detector, annotation, search and review of events. The main drawback of the system is it has filters which will work only for road surveillance. And it uses SIFT to search and store frames which have anomalies in it. But the main problem with searching by SIFT descriptors is the high dimensionality of the feature space, which is also why fast tree structures cannot be used.

Jerome Meessen et al. [3] proposed a system which focused on retrieval of particular configurations of objects the user is looking for as an example; one user may look for scenes presenting two pedestrians walking on the footpath, while the next user may wish to retrieve all scenes showing a white ambulance in a particular location. In this case the system has no prior knowledge of the target scene that is of interest to the user. The system adapts itself to the user needs which in turn are achieved by interactive learning known as active learning. The interactive interface allows the simultaneous visualization of both the classifier best results and the examples suggested by the active learning. The main drawback of this system is that the active learning doesn't always take into consideration the positive examples. After some phases the active learning may suggest some ambiguous examples which may not be relevant with the user feedback.

In year 2013, Huang Xuan et al. [4] proposed a surveillance video retrieval method based on detection of moving objects. They stored the video feeds into the database first and then applied various methods for moving object detection on the stored videos. They have detected moving objects via movement detection algorithm based on codebook, and realized automatic segmentation of surveillance video via calculating present Frame Movement Amount based on the foreground and background information in movement detection. They have used combination of SIFT and color feature for retrieval. The main drawback of this system is real time processing cannot be done and hence it requires more memory space and it is time consuming.

A framework for a video database system is explained by Yan Yang et al. [6] in year 2009, with tagging structure; which provided the ability to perform automatic search and retrieval for surveillance videos simultaneously. The video source can be any image grabber device such as a web camera. The processed frame stream is stored in a database while monitoring. In this method, tagging is the most important function where tags are stored in the database after processing. Then input SQL queries are fired on the database to find matching tags, retrieving all the related frames as output. This saves the time of searching the needed information. The biggest drawback of this system in extending it to real world applications is the data size. As the size of data to be processed increased, the performance of the system decreased.

Mahapatra et al. [7], proposed a method consisting of four main steps; Moving Object Detection, Feature Extraction, Feature Aggregation, and Human Contour Detection. The first step helps to detect all the foreground objects in an image sequence or video by background modeling. In the second step, from the silhouettes, contours of the foreground objects are extracted. In the third part; the three extracted features are aggregated into single feature vector using Fuzzy Inference System. Then in the last part; the contour is recognized, if the aggregated feature matches to a class of template database.

G. Medioni et al. [8], presented a system which takes as input a video stream obtained from an airborne moving platform and produces an analysis of the behavior of the moving objects in the scene. This system relies on two modular blocks: the first one detects and tracks moving regions in the sequence. It uses a set of features at multiple scales to stabilize the image sequence, that is, to compensate for the motion of observe, then extracts regions with residual motion and uses an attribute graph representation to infer their trajectories. The second module takes as input these trajectories, together with user-provided information to instantiate likely scenarios.

László Havasi et al. [9], introduced a video surveillance and event detection framework and application for semi-supervised surveillance use. The aim of this system is to generalize the interoperability, compatibility and legality in camera surveillance systems. The system's intended use is in automatic mode on camera feeds that are not actively watched by surveillance personnel, and should raise alarms when unusual events occur. They have presented the current 3 detector filters, and the extendable modular interface. Filters include local and global unusual motion detectors, left/stolen object detector, motion detector, tampering/failure detector, etc. for surveillance.

Kalpana Thakre et al. [10], proposed an effective CBVR system based on dominant features like motion,color and edge. The video clips present in the database are segmented into shots before feature extraction. The system works in three stages: shot segmentation, feature extraction and retrieving similar video clips for the given query clip. In feature extraction firstly motion features are extracted using squared Euclidean distance then colour features using colour quantization. The edge density feature is extracted for objects present in video clips. When the video clip is queried, the system retrieves a given number of video clips from the database that are similar to the query clip.The retrieval is performed based on the Latent Semantic Indexing, which measures the similarity between the database video clips and the query clip. The system is effective because of the features that are proposed to extract from any video clip. Those features have the capability of distinguishing video clips.

In year 2009, A. Singh et al. [13], presented an abandoned object detection system based on a simplistic and intuitive mathematical model which works efficiently at QVGA resolution which is the industry standard for most CCTV cameras. The overall system is modular in nature and consists of four disparate blocks with each block acting as a discreet processing unit. The 4 blocks are: Data extraction and conversion unit; Background subtraction module; still object tracking and occlusion detection block and Alarm raising and display of result unit.

Following section describes the proposed system architecture.

III. PROPOSED SYSTEM

Proposed system is used for detection of motion using a static camera working with its live feed and raising an alarm to notify the user. This strategy can be used in prevention and detection of crime as the system raises alarm for the unwanted motion. For example, in home surveillance if we set up the webcam and leave the house and if any motion is detected the owner will be notified immediately.

As shown in fig 1, as the authorized person logs in to the system, the camera will get live feed .The live feed will go under some pre-processing steps. The system will subtract the background frames and detect motion through centroid variance. After the motion is detected, the system will generate an alarm to warn the authorized individual.

A. *Live Camera Feed*

For taking the live feed from a camera, Jmyron library is used as an external library made specifically to capture live video feed from cameras.

B. *JMyron*

Myron is the cross-platform, cross-language, open source, video capture and computer vision plug-in. One core C++ object gets cross-compiled as a handful of high level language "wrapper" libraries. The wrapper for Java and Processing is called JMyron. The wrapper for Macromedia Director is called WebCamXtra. We can add other aspects like motion and color tracking. Some functions of Jmyron are:

1.Jmyron.start():It is used to start capturing the feed.
2.Jmyron.stop() :It is used to stop capturing the feed.
3.Jmyron.update():It is used to refresh image at every cycle.

C. *Pre-processing*

Pre-processing techniques are basically used in noise reduction and segmentation. This helps to view image more clearly. It is used to extract useful information from the total content. It includes processing an input image such as image or any video frame and the output of image will be either an image or set of parameters.

1) Grayscale: Value of each pixel is a single sample i.e intensity information. In grayscale, it is easy to calculate 8 bit information of each pixel than 24 pixel value in RGB color space.A grayscale digital image (black and white image) is an image in which the value of each pixel is a single sample, that is, it carries only intensity information. The intensity of a pixel is expressed within a given range between a Minimum(0) and a maximum(1). In proposed system grayscale will be helpful to differentiate between the intensity level for detecting the blobs.
Formula:
Averaging: R+G+B/3

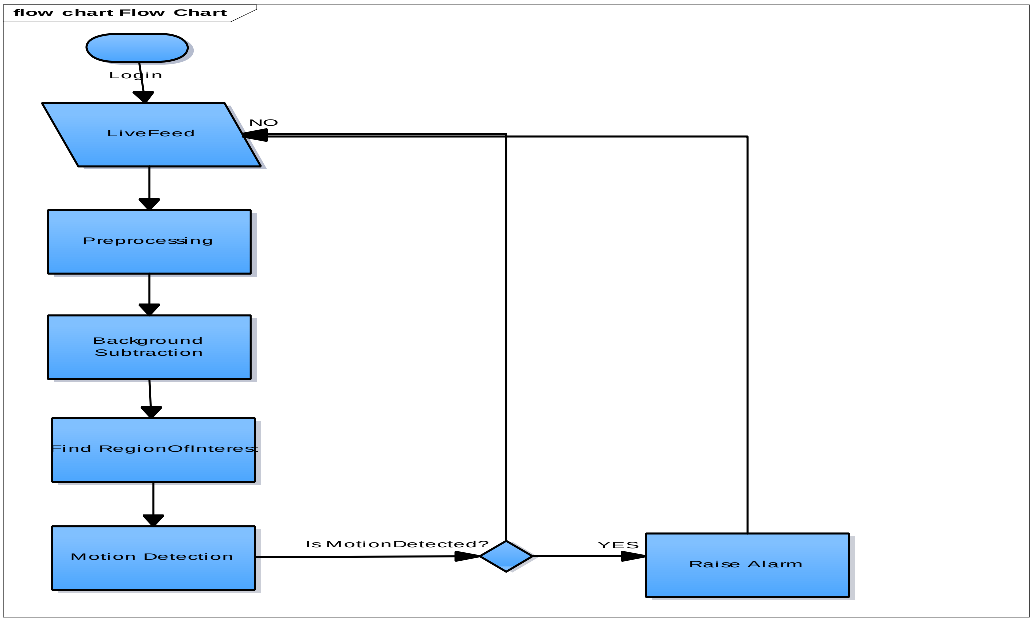

Fig 1 System Design

Fig 2 Grayscaling

2) Blurring: It is a technique used to remove image noise. In this method we simply reduce the edge content. Another name for blurring is smoothening. Blurring basically softens the image contours and reduces sharpness of the image. In proposed system, blurring will be used for reducing the noise which will soften the frame.

$$G(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{x^2}{2\sigma^2}}$$

D. Region of Interest

In this step we detect a region of interest through thresholding and blob detection methods. We use the hresholding and Blob detection algorithms to detect our region of interest which is further used for detection motion.

1) Thresholding:

Figure 3 Thresholding

Thresholding in the system is used to create binary images. Binary means images having only black and white colors. On setting a particular threshold, parts of the image can be made entirely white or entirely black depending on the requirement. For the proposed system, thresholding is used to identify objects and remove unwanted noise from the video feed. It can be applied after or before background subtraction. Different types of thresholding methods are listed below.

In year 2004, T. Sezgin and Sankur [5] categorize thresholding methods into the following six groups based on the information the algorithm manipulates:

- Histogram shape-based methods, where, for example, the peaks, valleys and curvatures of the smoothed histogram are analyzed
- Clustering-based methods, where the gray-level samples are clustered in two parts as background and foreground (object), or alternately are modeled as a mixture of two Gaussians
- Entropy-based methods result in algorithms that use the entropy of the foreground and background regions, the cross-entropy between the original and binarized image, etc.
- Object Attribute-based methods search a measure of similarity between the gray-level and the binarized images, such as fuzzy shape similarity, edge coincidence, etc.
- Spatial methods [that] use higher-order probability distribution and/or correlation between pixels
- Local methods adapt the threshold value on each pixel to the local image characteristics. In these methods, a different T is selected for each pixel in the image.

Our system works on the image and processes it to convert it to grayscale. Using threshold values we can isolate the objects. The algorithm is:

1) Traverse the entire input image/frame
2) Read individual pixel color value (24-bit) and convert it to grayscale.
3) Calculate the binary output pixel value (black or white) based on current threshold.
4) Store new value at same location in output image/frame.

Threshold can be calculated as:

If f (x, y) > T then f (x, y) = 0 else f (x, y) = 255

Where,

T = Threshold value.

2) Blob Detection: Blob detection refers to mathematical methods that are aimed at detecting regions in a digital image that differ in properties, such as brightness or color, compared to areas surrounding those regions. Informally, a blob is a region of a digital image in which some properties are constant or vary within a prescribed range of values; all the points in a blob can be considered in some sense to be similar to each other. Given some property of interest expressed as a function of position on the digital image, there are two main classes of blob detectors: (i) differential methods, which are based on derivatives of the function with respect to position, and (ii) methods based on local extreme, which are based on finding the local maxima and minima of the function.

For the proposed system, we will be using blob detection methods for detection of areas of interest. It will be the first step in detecting motion. The method used for blob detection works with pixel intensity tracking. The local minima and maxima of the image are calculated and image is converted to a grayscale image. Let us consider the case of detecting bright grey-level blobs and let the notation "higher neighbor" stand for "neighboring pixel that has a higher grey-level value". Then, at any stage, the algorithm (carried out in decreasing order of intensity values) is based on the following classification rules:

1. If a region has no higher neighbour, then it is a local maximum and will be the seed of a blob.

2. Else, if it has at least one higher neighbour, which is background, then it cannot be part of any blob and must be background.

3. Else, if it has more than one higher neighbour and if those higher neighbours are parts of different blobs, then it cannot be a part of any blob, and must be background.

4. Else, it has one or more higher neighbours, which are all parts of the same blob. Then, it must also be a part of that blob

3) Motion Detection: Motion detection is an upcoming field in science. Devices like Kinect use motion capture for superior gaming experiences. In our system, however, motion is used to detect an intruder and will be used to raise an alarm. Motion as a feature is a very vast concept. We will be taking the basic aspect of just detection and capture. Like [7] have mentioned in their paper, maximum algorithms for background modeling, work on gray scale video, after converting the RGB video frames to gray. Hence after our pre processing steps, we will try and detect a moving object through Centroid of the object.

The system proposes a method for motion detection where motion is detected through a method called Centroid Variance. This Method works with the detection of a blob and then taking out properties of this blob like area and centroid. These properties are then used to detect motion through variance. The centroid of the blob is calculated in every frame and then subtracted from the previous frame. Given a threshold value, if the value obtained from each comparison is more, motion is said to be present.

Next section will explain the basic UML designs of the system .

IV. UML Designs

A. Class Diagram

Class diagram gives a static view of a system. It helps in visualizing the aspects of a system and describing the code. Proposed system gives four classes, where main class is application GUI which is related with all other classes. The class admin describes the function which will be performed by the user. Class alert is used to give format in which the alert will be raised.

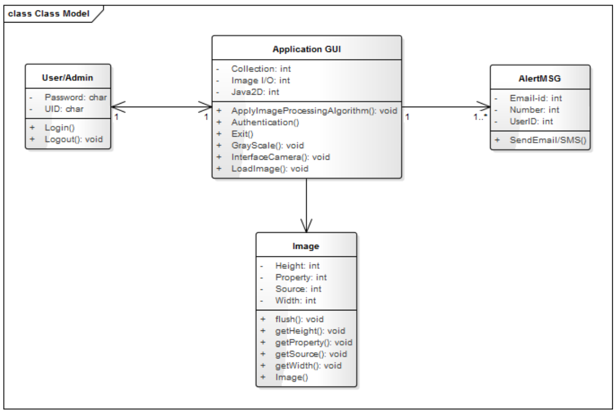

Figure 4 Class Diagram

B. Activity Diagram

Activity diagram gives the flow of operations in it. There are basically three timeline in the flow which are user, application GUI and camera interface. The basics operations are stated in diagram and the parallel operations which will be acting simultaneously. These are user login to application GUI and image loading through camera interface. After getting the camera feed, the background subtraction takes place, region of interest is obtained and if the motion is detected then alarm will be raised.

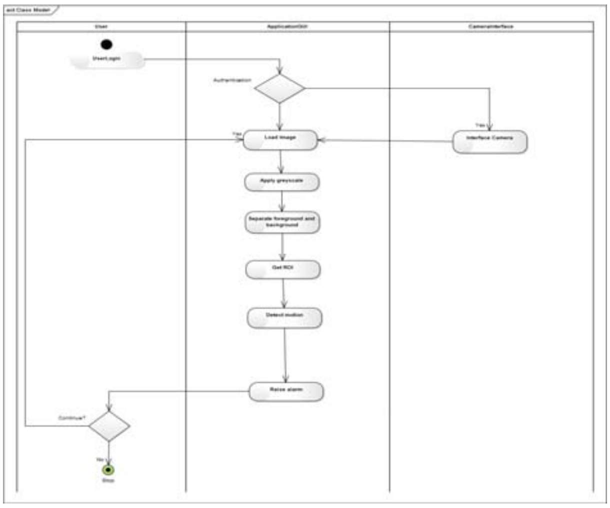

Fig. 5 Activity Diagram

C. State diagram:

State diagrams are used to give an abstract description of the behaviour of the system. Initially the system is in 'Idle' state. After starting the system, user 'Login' page is display and the user is authenticated. After successful login the user gets into 'Waiting state'. During this 'Waiting state' the image is loaded, reprocessed and algorithms are applied to detect the motion. While performing these tasks the states changes from waiting to 'InProgress'. If any unwanted motion is detected, then an alarm is raised and states changes to 'Complete'. If any process is aborted then the state changes from 'InProgress' to 'Cancel'. User logs out of the system once the work is completed.

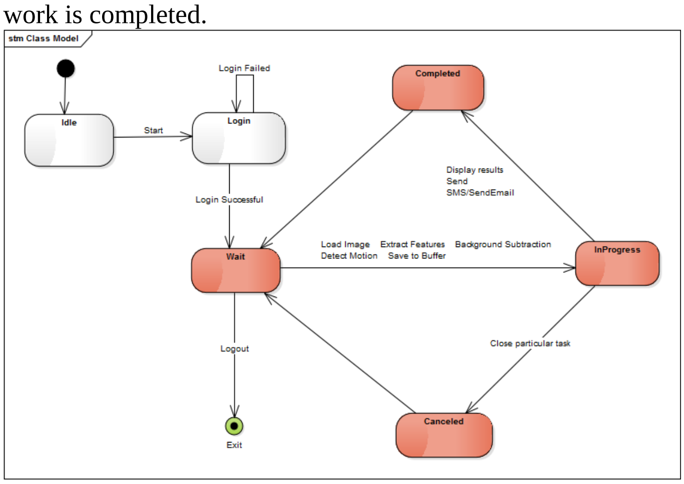

Figure 6 State Diagram

CONCLUSION

The proposed system will work to aid crime detection specifically in locations where the camera will be fixed. Surveillance can be interactive as the user is alerted about the motion detected in real-time. The system with the help of image and video processing techniques will work on specific key frames of surveillance footage and detect anomalies. All methods in the system are used because they will give the optimum results. Since the image is processed first with techniques like grayscaling and smoothening, all redundant noise, brightness, hazy pictures can be turned into ideal images. Hence no matter where the camera is situated, it won't affect the image quality. The system retrieves data in real time and hence no database is required for storage. The anomalous frame on capture gets stored automatically for further reference. Drawbacks of the system are that it will not work very well if the camera is moving. Applications of the system could be, Home Surveillance, Bank locker surveillance, Shop surveillance etc. Further enhancements can be made to the system where the user will be able to view the live feed from a remote location.

ACKNOWLEDGMENT

We would like to express our gratitude to all those who made it possible for us to complete this paper. First of all, we would like to thank Dr. K. S. Thakare for her help, support, encouragement, teaching and supervision. Her academic advice and ideas have played an important role in the work proposed in our paper. we are extremely grateful to our respected H.O.D (IT department) Dr. N.J. Uke for providing all facilities and every help for smooth progress. We would also like to thank all staff members of IT department for their timely help and inspiration.

REFERENCES

[1] J´erˆomeMeessen, MatthieuCoulanges, Xavier Desurmont, and Jean-Fran¸coisDelaigle, "Content Based Retrieval of Video Surveillance Scenes", Avenue Copernic 1, B-7000 Mons, Belgium.

[2] Levente Kovacs1, Zoltan Szlavik1, CsabaBenedek, Laszlo Havasi, IstvanPetras, David Losteiner1, AkosUtasi, Attila Licsar, Laszlo Czuni2, and Tamas Sziranyi1, "Video Surveillance Framework for Crime Prevention and Event Indexing", Distributed Events Analysis Research Group,Hungary.

[3] Jerome Meessen, Christophe De Vleeschouwer and Benoit M. Macq, "Flexible user interface for efficient content-based video surveillance retrieval: design and evaluation", ResearchGate, January 2009.

[4] Xuan, Fernando Croquer and Wenhua Zeng, "Surveillance Video Retrieval Based on Moving Objects Detection" , Journal of multimedia, Vol. 8, No. 6, December 2013.

[5] Mehmet Sezgin and Bulent Sankur, "Survey over image thresholding techniques and quantitative performance evaluation", Journal of Electronic Imaging 13(1), 146–165 (January 2004).

[6] Yan Yang,Brian C Lovell,Farhad Dadgostar , "Content-Based Video Retrieval (CBVR) System for CCTV Surveillance Videos", 978-0-7695-3866-2/09 $26.00 © 2009 IEEE DOI 10.1109/DICTA.2009.36

[7] Ansuman Mahapatra, Tusar Kanti Mishra, Pankaj K Sa, and Banshidhar Majhi "Background Subtraction and Human Detection in Outdoor Videos using Fuzzy Logic".

[8] G. Medioni, I. Cohen, F. Bremond, S. Hongeng, R. Nevatia, "Event detection and analysisfrom video streams", IEEE Trans. on Pattern Analysis and Machine Intelligence, vol. 23, no. 8,pp. 873 – 889, 2001.

[9] László Havasi, Zoltán Szlávik, Levente Kovács, Ákos Utasi, Csaba Benedek and Tamás Szirányi, "Validate Privacy Constraints in Surveillance Systems", Post-proceedings of the 2nd International Conference on ICT Solutions for Justice ,ICT4Justice 2009.

[10] KalpanaThakre, Archana Rajurkar and RamchandraManthalkar ,"An Effective CBVR System Based On Motion, Quantized Color And Edge Density Features", International Journal of Computer Science & Information Technology (IJCSIT), Vol 3, No 2, April 2011.

[11] Kalpana S Thakre,Archana M Rajurkar, R RManthalkar, "Content Based Video Retrieval using Latent Semantic Indexing and Color,Motion and Edge Feaures",International Journal of Computer Applications, Vol 54, No 12, September 2012.

[12] K.S.Thakre, A.M.Rajurkar, R.R.Manthalkar, "Video Partitioning and Secured Keyframe Extraction of MPEG video", International Conference on Information Security and Privacy.

[13] A. Singh, S. Sawan, M. Hanmandlu V.K. Madasu, B.C. Lovell, "An abandoned objectdetection system based on dual background segmentation", 2009 Advanced Video and Signal Based Surveillance.