WALCHAND COLLEGE OF ENGINEERING,SANGLI

(An Autonomous Institute)

Project Report

On

"SUSPICIOUS ACTIVITY DETECTION THROUGH VIDEO SURVEILLANCE SYSTEM"

Submitted in fulfilment of the requirement

for the award of the Degree

B. TECH

in

Information Technology

Submitted by

1. Aishwarya Adhiraj Gunde 2012BIT063
2. Shital Dnyanoba Jadhav 2012BIT018
3. Radhika Sujit Tiwari 2012BIT062

Under the guidance of

Mr. P. K. Kharat

DEPARTMENT OF INFORMATION TECHNOLOGY,

WALCHAND COLLEGE OF ENGINEERING,

SANGLI

CERTIFICATE

This is to certify that the project entitled

"SUSPICIOUS ACTIVITY DETECTION THROUGH VIDEO SURVEILLANCE SYSTEM"

Submitted by

1. Aishwarya Adhiraj Gunde 2012BIT063
2. Shital Dnyanoba Jadhav 2012BIT018
3. Radhika Sujit Tiwari 2012BIT062

3.

In the fulfillment for the award of Degree

BACHELOR OF TECHNOLOGY

IN

INFORMATION TECHNOLOGY

Is a bonafide record of their own work performed by them in the Final Year of Information Technology as specified in the curriculum prescribed by

WALCHAND COLLEGE OF ENGINEERING, SANGLI

Academic Year

[2015-2016]

Mr. P. K. Kharat	Dr.S.P.Sonavane
(Project Guide)	(H.O.D I.T.)
Date:	Place: WCE, Sangli

DECLARATION

We, the undersigned declare that the project entitled “Suspicious activity detection through video surveillence system” is the record of authentic work carried out by me during Fourth Year B.Tech in Information Technology as specified in the curriculum prescribed by Walchand College of Engineering, Sangli during the academic year 2015-2016.

Declaration by,

Name of Student	Exam Seat No.	Sign
1. Aishwarya Adhiraj Gunde	2012BIT063	
2. Shital Dnyanoba Jadhav	2012BIT018	
3. Radhika Sujit Tiwari	2012BIT062	

4.

Date:

Place: WCE, Sangli

Acknowledgement

Today on completion of this project dissertation, the persons we need to thank the most who have helped us throughout the making of this project dissertation and without whose help the project would not have seen the light of the day.

Primarily, We submit our gratitude and sincere thanks to our guide Mr. P. K. Kharat Sir, for his constant motivation and support during the course of the work in the last one year. We truly appreciate and value their esteemed guidance and encouragement from the beginning to the end of this project. We would also like to thank our HOD, Prof. Shefali Sonawane who encouraged us and created a healthy environment for all of us to learn in best possible way. We would like to thank all the staff members of our college and technicians for their help in making this project a successful one.

Name Of Student	Exam Seat No.	Sign
1. Aishwarya Adhiraj Gunde	2012BIT063	
2. Shital Dnyanoba Jadhav	2012BIT018	
3. Radhika Sujit Tiwari	2012BIT062	

Abstract

Video Surveillance systems are playing vital role in ensuring the security at various public places like bus stops, railway stations, shopping malls, Airports, etc. Currently, the intelligent video security surveillance systems are largely divided into system through image analysis and system based on location recognition applied to ubiquitous sensor network technology. Implementation of an effective surveillance system is a challenging task. In practice, a large number of CCTV cameras are installed to prevent illegal and unacceptable activities where a human operator observes different camera views and identifies various alarming cases [7]. Suspicious activity recognition helps to prevent from threats and identify the causes after threat. Existing semi-automatic approaches depends on human intervention to detect the uncommon activities and suspicious behavior from video context. Due to these limitations they become non-intelligence, very slow and need more human observers. This framework uses location of IT Department and context (foreground) change information for suspicious activity detection. The proposed work implements various existing techniques such as background subtraction, image enhancement techniques, feature extraction algorithms for processing and then constructs an algorithm based on them to detect missing objects.

Contents

1 Introduction **8**
- 1.1 Problem Statement . . . 9

2 Literature Survey **9**
- 2.1 Moving Object Detection: Review of Recent Research Trends . . 9
- 2.2 Evaluation of Background Subtraction Algorithms for Object Extraction . . . 10
- 2.3 Conception and Development of a Video Surveillance System for Detecting, Tracking and Profile Analysis of a Person . . . 10
- 2.4 Real-Time Semantics-Based Detection of Suspicious Activities in Public Spaces . . . 11
- 2.5 A Context-Based Approach for Detecting Suspicious Behaviors . 12
- 2.6 Sudden Event Recognition: A Survey . . . 12
- 2.7 Utility based Decision Support Engine for camera view selection in Multimedia Surveillance Systems . . . 12
- 2.8 From Smart Camera to SmartHub: Embracing Cloud for Video Surveillance . . . 13

3 Significance /Relevance **14**

4 Objectives **14**

5 Flow Diagram **15**
- 5.1 Frame Extraction: . . . 15
- 5.2 Elimination Of Redundant Frames: . . . 16
- 5.3 Background Substraction: . . . 16
- 5.4 Human Motion Detection: . . . 17
- 5.5 Object Detection: . . . 17
- 5.6 Counting Objects: . . . 18
- 5.7 Alarm Generation: . . . 18

6 Technical Highlights **19**
- 6.1 CCTV: CP PLUS . . . 19
- 6.2 MATLAB R2013a . . . 19
- 6.3 Kalman Filter: Human Motion Detection . . . 20
- 6.4 Feature Extraction Algorithms . . . 21
  - 6.4.1 FAST: . . . 21
  - 6.4.2 SURF: . . . 21
  - 6.4.3 HARRIS: . . . 21
- 6.5 Image Enhancement Algorithms . . . 22
  - 6.5.1 Unsharp Mask Filtering: . . . 22
  - 6.5.2 Histogram Equalization: . . . 22
  - 6.5.3 Wiener Filtering: . . . 22

7 Experiments **23**

8 Results **25**
8.1 Image of GUI: 25
8.2 Image of scene: 26
8.3 Image of Object1 (Mouse): 27
8.4 Matched features of Mouse: 28
8.5 Detection of Mouse: 29
8.6 Image of Object2 (Keyboard): 30
8.7 Matched features of Keyboard : 31
8.8 Detection of Keyboard: 32
8.9 Detection of missing CPU: 33

9 Conclusion **34**

10 Future Work **35**

11 References **36**

1 Introduction

The name MATLAB stands for MATrix LABoratory.MATLAB is a high-performance language for technical computing. It integrates computation, visualization, and programming environment. Furthermore, MATLAB is a modern programming language environment: it has sophisticated data structures, contains built-in editing and debugging tools, and supports object-oriented programming. These factors make MATLAB an excellent tool for teaching and research. MATLAB has many advantages compared to conventional computer languages (e.g., C, FORTRAN) for solving technical problems. MATLAB is an interactive system whose basic data element is an array that does not require dimensioning.It has powerful built-in routines that enable a very wide variety of computations. It also has easy to use graphics commands that make the visualization of results immediately available. Specific applications are collected in packages referred to as toolbox. There are toolboxes for signal processing, symbolic computation, control theory, simulation, optimization, and several other fields of applied science and engineering.

1.1 Problem Statement

To detect suspicious activities in particular area that is under CCTV surveillance using video processing tools .

2 Literature Survey

2.1 Moving Object Detection: Review of Recent Research Trends

The foremost part of video processing for activity detection lies in the determining the motion of objects that are part of video. Hence, several approaches are being put forth to detect motion of objects and human tracking for video processing. For surveillance systems, it is important to consider the motion detection so that it could be further classified into normal and suspicious activities based on experimental results. Traditional Approaches for moving object detection can be broadly categorized into four forms as Background Subtraction, Frame Differencing, Temporal Differencing and Optical Flow [1]. Background subtraction works by initializing a background model, then difference between current frame and presumed background model is obtained by comparing each pixel of the current frame with assumed background model color map.[1] In case difference between colors is more than threshold, pixel is considered to be belonging to foreground. Performance of traditional background subtraction method mainly gets affected when background is dynamic, illumination changes or in presence of shadow. [1]. Frame difference method identifies the presence of moving object by considering the difference between two consecutive frames [1]. The traditional approach makes use of image subtraction operator that obtains output image by subtracting second image frame from first image frame in corresponding consecutive frames. Frame differencing method lacks in obtaining the complete contour of the object as a result of which morphology operations are general used to obtain better results. [1] Temporal differencing method detects the moving target by employing pixel-wise difference method among two successive frames [1]. Traditional temporal difference method is flexible to dynamic changes in the scenes but results degrade when moving target moves slowly since due to minor difference between consecutive frames, object is lost. Optical flow approach of moving target detection is based on calculation of optical flow field of image (or video frame).Clustering is performed on the basis of the obtained optical flow distribution information obtained from the image (video frame). [1] This method allows obtaining complete knowledge about the movement of the object and is useful to determine moving target from the background. However, this method suffers from some of drawbacks like large quantity of calculations are required to obtain optical flow information and it is sensitivity to noise. [1]

2.2 Evaluation of Background Subtraction Algorithms for Object Extraction

Background Subtraction is one of the first approaches towards object detection in video processing. Thus, the technique is widely used in surveillance applications. Some drawbacks of background subtraction algorithm include variations in illumination due to presence of shadow or noise. A slight change in background frame can affect the foreground detection. Thus, the background subtraction algorithms need to be modified such that it could handle slight changes in the background without altering the expected foreground detection. Simple Background Subtraction (SBS) makes an initial assumption that the reference background is available, from which each time current frame is simply subtracted for the foreground/object extraction [2]. The Running Average (RA) method where rather than single reference background takes the average of all the incoming video frames to estimate the background rather than single reference background [2]. It has been observed in most background modeling techniques that they exploit sequence of linear evaluations to construct their backgrounds. Such sequence linear evaluation can be easily implemented in the DCT domain. Running Average algorithm uses such sequence of linear evaluations to estimate the background model. Thus this technique can be extended in the DCT domain called as RADCT algorithm to estimate the background. [2] In accordance with the pixel-based decision framework, the temporal statistics of pixels of the original video sequence is calculated by a new background subtraction method called Sigma Delta Estimation (SDE) method. [2] Simple Statistical Difference (SSD) method estimates the binary motion detection mask by finding mean and standard deviation values calculated with respect to each pixel of the previous video frames. Background image is produced from the collection of previous frames in the time interval (t0, tk-1) for each pixel calculating mean (xy) and standard deviation (xy). [2]

2.3 Conception and Development of a Video Surveillance System for Detecting, Tracking and Profile Analysis of a Person

Human body tracking is one of the basic requirements in video surveillance. Since the suspicious activities being carried out in a surveillance area can be by human intervention itself, it is necessary to analyze human motion. The majority of computer surveillance involves the monitoring of data like tracking people. Automated video surveillance is an essential research area in the business sector as well. [3] Mixture of Gaussian Model has been one of the most popular background subtraction techniques in computer vision because of its robustness to subtle illumination changes. Its bottleneck is its computational intensity because of the need to compute and update the Gaussian mixtures. [3]

2.4 Real-Time Semantics-Based Detection of Suspicious Activities in Public Spaces

Activity recognition and classification according to constraints are the core components of the surveillance application while dealing with suspicious activity detection. Thus, it is important to maintain the accuracy while classifying the activities into desired category. The use of computerized surveillance systems in large public areas such as metro stations and airports has been increasingly expanding. However, these systems are labor-intensive, and the monitoring personnel's prompt attention has been found to be improbable and to drop rapidly with time.[4] Thus the capacity to use surveillance systems to prevent crime rather than just for forensics has not yet been fully utilized. Thus, the function of an enhanced application for surveillance system is to make the required to flag prescribed suspicious behaviors when they happen. The behaviors of interest are defined in terms of motion features. To compute these, each frame is processed using a background subtraction method from which foreground blobs are obtained. Then, the spatial and temporal relations of these low-level blobs are calculated. Using these relations, high-level objects of interest that represent semantic entities are detected and tracked from frame to frame. [4] The single-object features are stored in historical sequence and used to make temporal inferences about trajectory and state. As well, the inter-object features between every combination of two objects is calculated and used for behavior recognition. The object classification includes different categories of depending upon the behaviors of interest. The behaviors addressed in [4] are abandoned luggage, loitering, fighting and fainting. Abandoned luggage detection has been widely researched in the literature. It is mostly addressed by using only background subtraction methods, without requiring other higher-level forms of reasoning such as object classification and tracking. However, this approach cannot discriminate between a stationary person and abandoned luggage. [4] To work around this issue, some researchers propose characterizing abandoned objects in term of a number of features such as color, edges, shape completeness, and histogram contrast. Still, none of the aforementioned papers addresses the issue of finding the object's owner. It would seem that this would be crucial, since a bag cannot be considered abandoned if the owner does not leave it, even if it has been stationary for a long time. [4] Loitering is useful for detecting a number of public transit situations such as drug dealing. As defined in [4], loitering is "the presence of an individual in an area for a period of time longer than a given time threshold". Fight detection poses a challenge due to the difficulty of characterizing the fighting activity. The most reliable means for defining fights is in terms of the frequency of object splitting and merging. [4] However, people standing still or walking together might split and merge due to some slight movements. Therefore, a 'minimal speed in different directions' condition should be added.[4] Perhaps the simplest and most widely used feature for detecting fainting people is the aspect ratio of the blob representing the person [4, 33-34]. Despite its popularity, tests in [4] have indicated that this approach fails to detect situations where people fall to the ground while being aligned with the camera or where camera nonlinearities are significant. Evaluation of behavior recognition is inherently difficult because of the high complexity of the actions and the presence of other confusing factors in the test scenarios.

2.5 A Context-Based Approach for Detecting Suspicious Behaviors

With the increasing use of CCTV for crime detection and safety, it has become necessary to ensure that the feeds from the surveillance are properly supervised. This leads to increased workload among the human operators and delay in response to the crime activity. Thus, a need for automated system is the demand for surveillance systems. Several research papers have proposed various alternatives regarding detection of abnormal or suspicious activities and need for automated surveillance systems. A context based approach implies to technique where suspicious activities are detected utilizing the contextual information. Two contextual features, namely, type of behavior and the commonality level of each type are extracted from long term observation. Then, a data stream model which treats the incoming data as a continuous stream of information is used to extract these features. [5] When making a decision whether or not an observed behavior is suspicious, one may rely on his/her contextual knowledge. The knowledge could be gathered from previous observations at either the same place or other places with similar context. [5] By using these contextual features, the system is able to learn not only from the training sets but also adapt itself to real-time video data.

2.6 Sudden Event Recognition: A Survey

Event detection has become an important research area in which many researchers have focused on classifying the event as either a normal or abnormal event. Consequently, abnormal event recognition becomes a necessity in surveillance systems to ensure safety and comfort. An abnormal event is typically assumed to be similar to the terms unusual, rare, atypical, surprising, suspicious, anomalous, irregular and outlying. Thus, a sudden event can be summarized as: (1) a subset of an abnormal event; (2) an event that occurred unexpectedly, abruptly and unintentionally that invokes an emergency situation; and (3) an event that is detected in a few numbers of frames, which requires a fast response to mitigate the posterior risks. [6]

2.7 Utility based Decision Support Engine for camera view selection in Multimedia Surveillance Systems

It has been reported that CCTV operators suffer from boredom since in surveillance places nothing actually happens. The human operator may not be able to pay attention to all the camera views at the same time. As the current surveillance system suffers from lack of supervision, there should be some constraints by which the human operator could judge the most appropriate view which needs more attention. [7] Presents a DSE that selects and schedules appropriate camera views for the operator so that he can give full attention to the selected views and observe them carefully. For this purpose, data gathered from sensors are processed, alarms are detected, classified and ordered according to threat severity. Aside from these, a set of scenarios in the context of surveillance application is defined to identify behavioral patterns. The aim of this procedure is to represent and recognize complex activities under surveillance from simple behaviors at lower levels. [7] The proposed method selects and schedules the most appropriate and relevant camera views for human operators. The entire process composed of several intermediate processing units. The first step is to capture data from different sensors especially from surveillance cameras. Two types of events are considered namely significant and insignificant events. Insignificant event refers to event with low utility value and can be ignored. Thus, the decision for selecting camera views for human operators does not depend on the camera sources of the insignificant events. The significant events have comparatively high utility value and are considered as candidates for the view selection. The human operator can also check the logs when necessary.

2.8 From Smart Camera to SmartHub: Embracing Cloud for Video Surveillance

While focusing on the techniques and algorithms for video processing, it is equally important to consider the video quality and device that is to be used for monitoring. Despite great progress in terms of research, smart cameras have not seen enough success in commercial systems and real installations. [8] investigates the reasons behind the restricted use of smart cameras even though they reduce the workload and send only the abstract for further processing. It is found that smart cameras are effective only for sparse camera networks where multisensor information fusion is minimal, such as a highway traffic monitoring systems, but they are not suitable for applications requiring the assimilation of data from multiple cameras. [8] It discusses the issues related with smart cameras such as limited opportunity for information fusion. It proposes a SmartHub system where all the tasks related to video processing such as tracking and regconition is is carried out by SmartHub cloud entity. SmartHub fuses visuals from multiple cameras and provides a set of services to the central unit such as object detection, face detection, and tracking. The central unit can query SmartHub to receive continuous information (video streams) or event information in terms of detected objects. Because the information fusion takes place at SmartHub, it does not need to send video from all cameras to the central unit but only themost informative view. [8] It states the merits of SmartHub systems which includes Storage Scalability and Reduced Processing Repetition Lower per Sensor Cost.

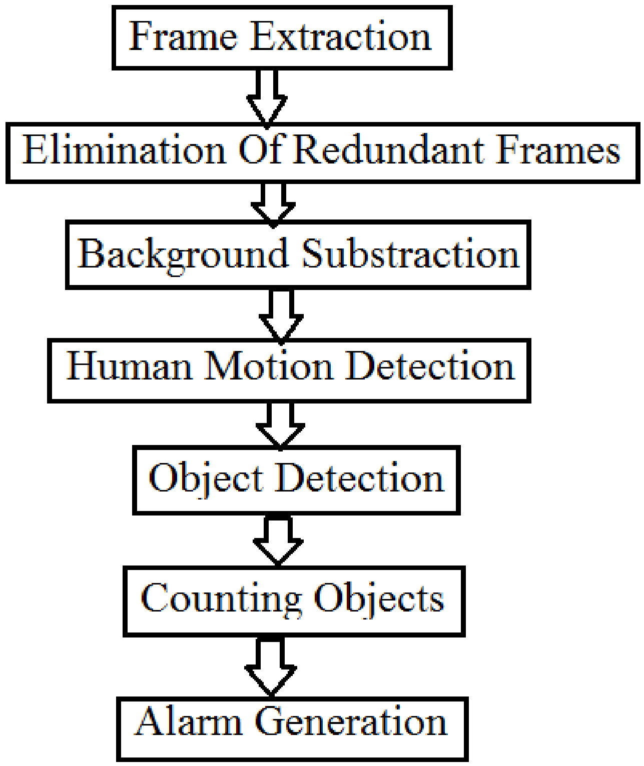
3 Significance /Relevance

Several important research questions remain to be addressed before we can rely upon video surveillance as an active tool for crime prevention, crime resolution, and crime protection. Video surveillance is increasing signi_cance approach as organizations seek to safe guard physical and capital assets. At the same time, the necessity to observe more people, places, and things coupled with a desire to pull out more useful information. In a normal retail environment, it is common for video storage costs to exceed 50% of the entire surveillance system cost. This places an incredible financial burden on retailers, especially during these tough economic times. In response, many retailers are turning to "record-on-motion" devices that have the potential to save storage, but can leave costly gaps in visibility. Thus, there is need of management and automated processing to prevent excess space usage. These are some crucial aspects of video surveillance that are needed to be addressed.

4 Objectives

1. Management and analysis of huge amount of video data.
2. Study and comparisions of algorithms for detection purpose.
3. Study of video processing for detecting suspicious activities in IT Department through CCTV surveillance.

5 Flow Diagram



5.1 Frame Extraction:

Video is the recording, reproducing, or broadcasting of moving visual images. Thus, in order to apply any further processing technique, we need to extract frames from video. This can be done using MATLAB functions, so as to implement the image processing techniques as per the requirement. The implemented MATLAB code reads the video using the VideoReader() function in terms of image frames. Then the imwrite() function stores the frames on PC. Using get() function, it is able to identify the features of video file and that the frames were extracted at the rate of 25fps.

5.2 Elimination Of Redundant Frames:

Elimination of redundant frames was of great significance for our project since we have to work on large video files having the aforementioned frame extraction rate. There were too much redundant frames in video. For making space utilization, we removed redundant frames. To reduce redundant frames we firstly convert them into grayscale, the frames having intensity below threshold values are deleted. Hence, efficiency of computation and space optimization were achieved.

5.3 Background Substraction:

Background Subtraction Method is considered to be one of the most reliable method for moving object detection. Background subtraction works by initializing a background model, then difference between current frame and presumed background model is obtained by comparing each pixel of the current frame with assumed background model color map. In case difference between colors is more than threshold, pixel is considered to be belonging to foreground [6]. Performance of traditional background subtraction method mainly gets affected when background is dynamic, illumination changes or in presence of shadow. Numerous methods have been developed so forth to upgrade background subtraction method and overcome its drawbacks. Different methods of background subtraction as reviewed by Massimo Piccardi et al. [7] are: Concurrence of image variations, Eigen backgrounds, Mixture of Gaussians, Kernel density estimation (KDE), Running Gaussian average, Sequential KD approximation and Temporal median filter.

The proposed work firstly separates foreground image pixels from background image pixels by learning the background from the presumed model, using Gaussian mixture strategy. Subsequently, movement of the object is traced within the frame by blob detection. To recognize area of the blob and to calculate detected region and its centroid, blob detection makes use of contrast in binary image. The pixels identified by Gaussian mixture model are grouped into disconnected classes through contour detection algorithm. The disconnected classes and its surrounding contours are tagged as candidate blob. In order to reduce false detection, the small candidate blobs are removed. The positions of contour blob are compared by using k-means clustering that identifies centre of clusters to detect vehicles in each region. Some morphological operations are applied for elimination of noise. The blob analysis puts a rectangular box around the potential objects. Finally, the counting of vehicles within the baseline is done based on the identified blob regions.

5.4 Human Motion Detection:

We used the recursive estimator algorithm of Kalman Filter to detect human motion in the video. A rectangular boundary box was used to indicate any kind of multiple human motions simultaneously taking place in the CCTV footage. As we had to detect suspicious activities being carried out, human motion detection was not enough as this real-time detection algorithm tracked all kinds of motion, be it normal or anomaly. Thus, it was necessary to look for object matching algorithms that could trace the objects in the video and some sort of analysis could be done in case of their absence. All these further led to the study of object detection.

5.5 Object Detection:

We used algorithm for detecting a specific object based on finding point correspondences between the reference and the target image. It can detect objects despite a scale change or in-plane rotation. It is also robust to small amount of out-of-plane rotation and occlusion.

This method of object detection works best for objects that exhibit non-repeating texture patterns, which give rise to unique feature matches. This technique is not likely to work well for uniformly-colored objects, or for objects containing repeating patterns. Note that this algorithm is designed for detecting a specific object, for example, the CPU in the reference image, rather than any CPU. For detecting objects of a particular category, such as people or faces, see vision.PeopleDetector and vision.CascadeObjectDetector.

Step 1: Read Images

Read the reference image containing the object of interest.

Step 2: Detect Feature Points

Detect feature points in both images.

Step 3: Extract Feature Descriptors

Extract feature descriptors at the interest points in both images.

Step 4: Find Putative Point Matches

Match the features using their descriptors.

Step 5: Locate the Object in the Scene Using Putative Matches

estimateGeometricTransform calculates the transformation relating the matched points, while eliminating outliers.

This transformation allows us to localize the object in the scene.

Step 6: Detect Another Object

Detect a second object by using the same steps as before.

Read an image containing the second object of interest.

5.6 Counting Objects:

We count the number of objects in lab. When count of objects is less than actual then it seems that some object is missing in the lab. We detected the missing object and represent the missing object in colour .

5.7 Alarm Generation:

As soon as object is missing in lab alarm is generated using sound() function. Also email is send to the receipient using sendmail() function.

6 Technical Highlights

6.1 CCTV: CP PLUS

Cp Plus is leading brand of CCTV camera in India. CP plus India has huge network of CCTV Camera distributors in India in Every state & city. CP Plus is providing different type of security camera such as IP Camera, Dome Camera, HD Camera Wireless Camera, and Bullet Camera. PTZ Camera, Wifi Camera, Hidden Camera, DVR, NVR, Video Door Phone, Biometric Attendance, VDP, Analog Camera, Night Vision Camera, Network Camera, Spy Camera, Cube Camera, Pinhole Camera, Outdoor Camera, Zoom Camera, Home Automation, Varifocal Camera, Surveillance Camera, Digital Video Recorder, Door Bell Camera, SMPS, Access Control, CP Plus Cable and Software.

This CCTV is used in our department's lab. It uses the Phase Alternating Line (PAL) video standard, which is a colour encoding system for analogue television used in broadcast television systems in most countries broadcasting at 25 frame per second. Other common colour encoding systems are NTSC and SECAM. It has (1024*768) pixel resolution.

6.2 MATLAB R2013a

MATLAB (MATrix LABoratory) is a multi-platform, data analysis, prototyping, and visualization tool with built-in support for matrices and matrix operations, rich graphics capabilities, and a friendly programming language and development environment. It offers programmers the ability to edit and interact with the main functions and their parameters, which leads to valuable time savings in the software development cycle. It has become very popular with Engineers, Scientists, and Researchers in both Industry and Academia due to many factors such as the availability of toolboxes containing specialized functions for many application areas, ranging from data acquisition to image processing.

6.3 Kalman Filter: Human Motion Detection

The Kalman Filter (named after its inventor, Rudolf E. Kalman) is an efficient recursive computational solution for tracking a time- dependent state vector with noisy equations of motion in real time by the least-squares method. It is used to separate signal from noise so as to optimally predict changes in a modeled system with time. The Kalman filter is a recursive estimator. This means that only estimated state from the previous time step and the current measurement are needed to compute the estimate for the current state.

The Kalman filter algorithm works as follows

Initialize $\hat{\mathbf{x}}_{0|0}$ and $\mathbf{P}_{0|0}$.

At each iteration $k = 1, \ldots, n$

Predict

Predicted (a priori) state estimate

$$\hat{\mathbf{x}}_{k|k-1} = \mathbf{F}_k \hat{\mathbf{x}}_{k-1|k-1} + \mathbf{B}_k \mathbf{u}_k$$

Predicted (a priori) estimate covariance

$$\mathbf{P}_{k|k-1} = \mathbf{F}_k \mathbf{P}_{k-1|k-1} \mathbf{F}_k^{\mathrm{T}} + \mathbf{Q}_k$$

Update

Innovation or measurement residual

$$\tilde{\mathbf{y}}_k = \mathbf{z}_k - \mathbf{H}_k \hat{\mathbf{x}}_{k|k-1}$$

Innovation (or residual) covariance

$$\mathbf{S}_k = \mathbf{H}_k \mathbf{P}_{k|k-1} \mathbf{H}_k^{\mathrm{T}} + \mathbf{R}_k$$

Optimal ***Kalman gain***

$$\mathbf{K}_k = \mathbf{P}_{k|k-1} \mathbf{H}_k^{\mathrm{T}} \mathbf{S}_k^{-1}$$

6.4 Feature Extraction Algorithms

6.4.1 FAST:

FAST (Features from Accelerated Segment Test) algorithm was proposed by Edward Rosten and Tom Drummond in their paper "Machine learning for high-speed corner detection" in 2006. It is a corner detection method, which could be used to extract feature points and later used to track and map objects in many computer vision tasks. The most promising advantage of FAST corner detector is its computational efficiency. Referring to its name, it is fast and indeed it is faster than many other well- known feature extraction methods. Moreover when machine learning method is applied, a better performance could be achieved which takes less time and computational resources. FAST corner detector is very suitable for real-time video processing application because of high-speed performance.

6.4.2 SURF:

In computer vision, Speeded Up Robust Features (SURF) is a local feature detector and descriptor that can be used for tasks such as object recognition or registration or classification or 3D reconstruction. It is partly inspired by the scale-invariant feature transform (SIFT) descriptor. The standard version of SURF is several times faster than SIFT and claimed by its authors to be more robust against different image transformations than SIFT. To detect interest points, SURF uses an integer approximation of the determinant of Hessian blob detector, which can be computed with 3 integer operations using a precomputed integral image. SURF descriptors can be used to locate and recognize objects, people or faces, to make 3D scenes, to track objects and to extract points of interest.

6.4.3 HARRIS:

Feature detection is a preprocessing step of several algorithms that rely on identifying characteristic points or interest points so as to make correspondences between images, recognize textures, categorize objects or build panoramas. The Harris corner detector algorithm relies on a central principle: at a corner, the image intensity will change largely in multiple directions. This can alternatively be formulated by examining the changes of intensity due to shifts in a local window. Around a corner point, the image intensity will change greatly when the window is shifted in an arbitrary direction. Following this intuition and through a clever decomposition, the Harris detector uses the second moment matrix as the basis of its corner decisions. The matrix has also been called the autocorrelation matrix and has values closely related to the derivatives of image intensity.

6.5 Image Enhancement Algorithms

6.5.1 Unsharp Mask Filtering:

Unsharp masking (USM) is an image sharpening technique, often available in digital image processing software. The "unsharp" of the name derives from the fact that the technique uses a blurred, or "unsharp", negative image to create a mask of the original image. The unsharped mask is then combined with the positive (original) image, creating an image that is less blurry than the original. In the photographic unsharp masking procedure, the amount of blurring can be controlled by changing the "softness" or "hardness" (from point source to fully diffuse) of the light source used for the initial unsharp mask exposure, while the strength of the effect can be controlled by changing the contrast and density (i.e., exposure and development) of the unsharp mask.

6.5.2 Histogram Equalization:

Histogram equalization is a method in image processing of contrast adjustment using the image's histogram. Adaptive histogram equalization (AHE) is a computer image processing technique used to improve contrast in images. It differs from ordinary histogram equalization in the respect that the adaptive method computes several histograms, each corresponding to a distinct section of the image, and uses them to redistribute the lightness values of the image. It is therefore suitable for improving the local contrast and enhancing the definitions of edges in each region of an image. However, AHE has a tendency to overamplify noise in relatively homogeneous regions of an image.

6.5.3 Wiener Filtering:

The Wiener filter as originally proposed by Norbert Wiener is a signal processing filter which uses knowledge of the statistical properties of both the signal and the noise to reconstruct an optimal estimate of the signal from a noisy one-dimensional time-ordered data stream. The generalized Wiener filter generalizes the same idea beyond the domain of one-dimensional time-ordered signal processing, with two- dimensional image processing being the most common application.

7 Experiments

The major objective of detecting suspicious activities in the department's lab through video surveillance has been accomplished by integrating the aforementioned algorithms. The entire work was divided into simpler and smaller tasks inclusive of frame extraction, elimination of redundant frames, background subtraction, human motion detection, object detection, image enhancement, counting objects and alert generation.

The primary step we carried out was about the study of image and video processing in MATLAB R2013a tool. As MATLAB works efficiently on images, frame extraction was fundamental for further analysis of video data. The digital data storage has always been a major challenge and in order to tackle the related problem, we eliminated some redundant frames. The image comparison in MATLAB is done based on its grayscale. To obtain difference between the images, we had to subtract the previous frame from the current frame and note its absolute difference. By carrying out some trial and error, we set a threshold of notable quantity, the value below which the frames get deleted.

After skipping the storage of redundant frames, we worked on the concept of background subtraction. Background Subtraction, also known as Foreground Detection, is a technique wherein an image's foreground is extracted for further processing. In this case, the regular and steady scene of the lab was considered as background image. Any change in this background frame, be it incoming and outgoing of crowd or displacement of already present objects was identified by subtracting the background image from the respective current frame.

Human Motion Detection was our next process to the background subtraction. Using the Kalman Filter algorithm, we detected multiple human motions in the set of frames of video. We were successful in indicating the human motion in real time video.

In order to detect suspicious activity, we first had to carry out one in the department lab. The major suspicious scenario we are focusing on is regarding thievery in the lab, for example, taking away Monitor or CPU from the lab. For this purpose, we had to detect objects in the frames. Here, objects are nothing else but Keyboard, Mouse, Monitor and CPU. Initially, we stored the images of set of these objects. We used various Extraction Feature Algorithms to match the object images with the frames of video. These algorithms work such that once the feature points are matched, the object is detected on the respective frame by creating a boundary box around the detected portion.

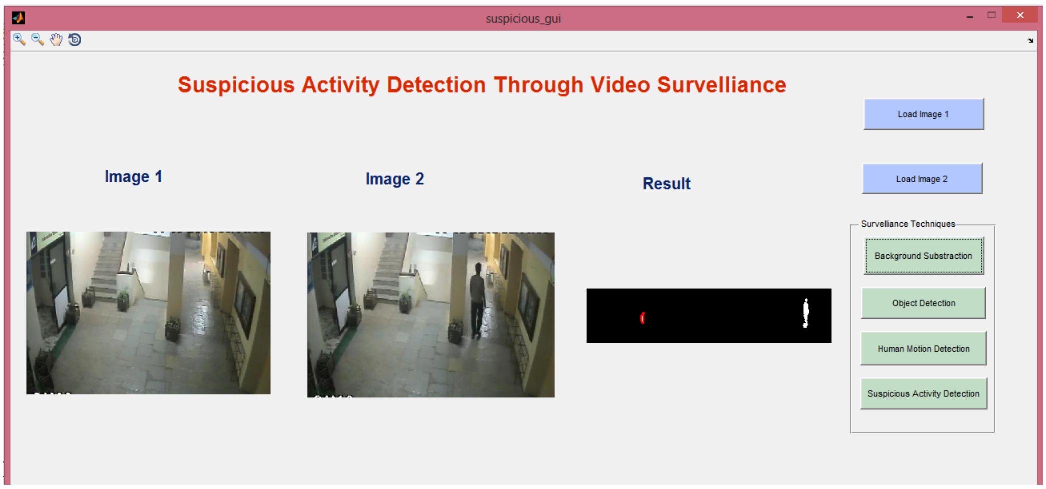
The extraction features algorithms we went through for object detection were SURF, Harris, FAST, MinEigen, etc. Another challenge we faced was concerning the resolution of CCTV (Closed Circuit Television) camera! Due to the angle and location of the camera, some of the objects such as mouse appeared very tiny. Consequently, the features could hardly be extracted. In spite of trial of various other extraction algorithms, Preprocessing Image Enhancement was the only solution we found out. MATLAB provides numerous techniques for the same such as initially blurring the image and later deblurring, and so on.

The major step in our project was to integrate the object detection algorithm with the counting objects algorithm for Detection of Suspicious Activities. The counting objects algorithm includes conversion of the image into BW image. Using the bwboundaries function for the converted image, we store the value of length returned by that function. This length gives the count of objects in the image. For the matching point's theory of the object image and frames of video, we included the estimateGeometricTransform function. This particular function returns 1 on the status of 'Not enough matching points found'. So, if it returns 1, then the image is cropped into the object image by applying defined cropping parameters. Next, counting objects algorithm was applied to both the images. If the count of the objects in the cropped image of the frame is less than the object image, then definitely some of the objects in the frame are missing. This indicates that may be Monitor, CPU, Keyboard or Mouse of the lab has gone missing from their respective position!

Once we were indicated that an object is missing, we had to find out which one is missing. Thus, to show the missing object in the frame, we used 'imfuse' function with falsecolor, blend, diff and montage methods that displayed the composite image of background frame and missing object image. Alarm Generation was the last task towards achieving our target. We used the sound MATLAB function in order to create some alarming noise once the count of object in frame was found to be less than count of object in the stored object image. Additionally, we used sendmail MATLAB function to generate alert emails to the respective department recipients or supervisors. By all this entire procedure, we could find out suspicious activity carried out and then generate alarming sound and alert messages to inform about the incident taken place.

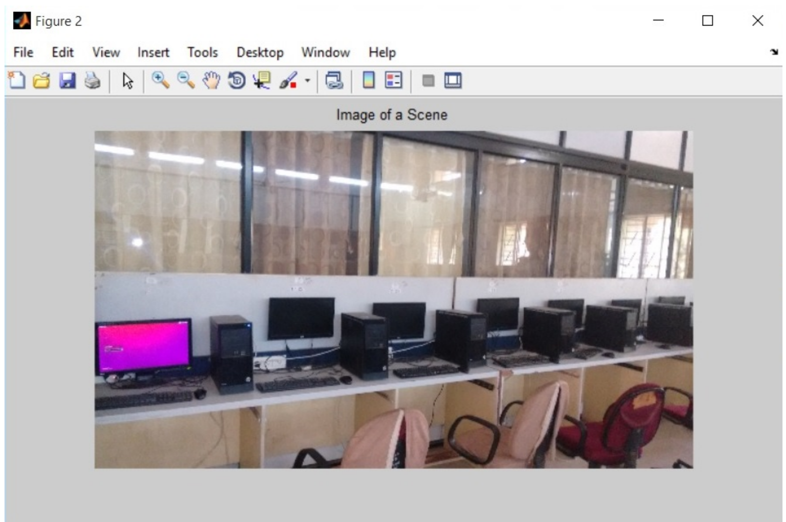
8 Results

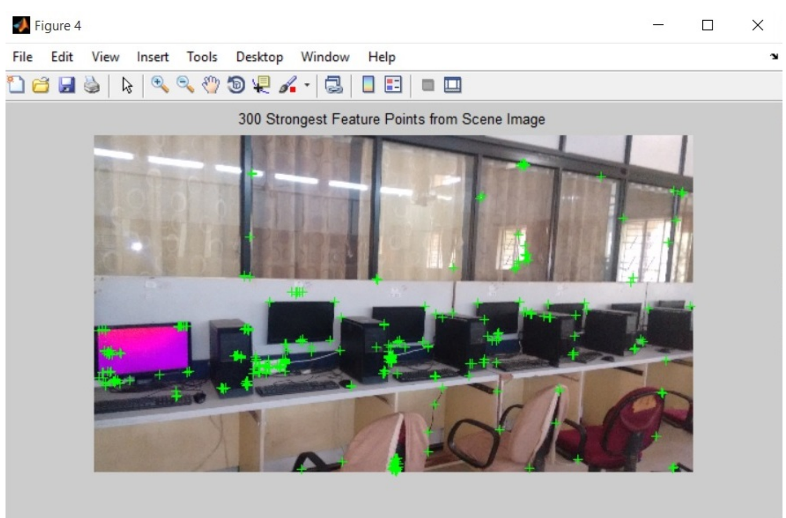
8.1 Image of GUI:



We created the Graphical User Interface (GUI) using GUIDE, which is the Graphical User Interface Devolopment Environment of MATLAB. GUIDE automatically generates the MATLAB code for constructing UI. We have added buttons, panels, axis to provide a user-friendly and easy to navigate look. Users can load images and perform the surveillance techniques which are provided in the panels. This figure shows the background substraction technique.

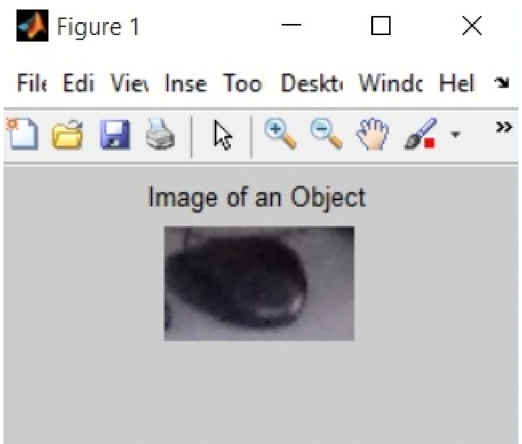
8.2 Image of scene:

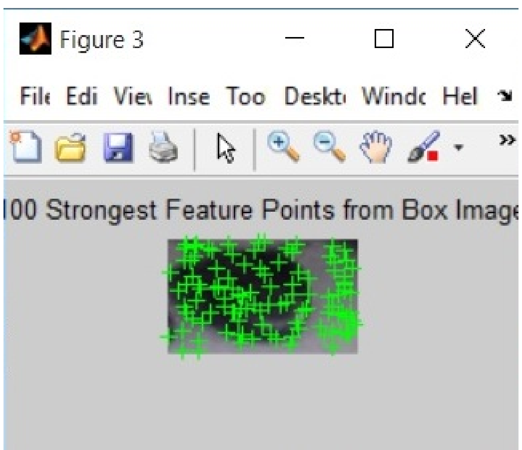




Original image of scene and the scene image with strongest feature points are shown in the figures. It is required to implement object detection algorithm.

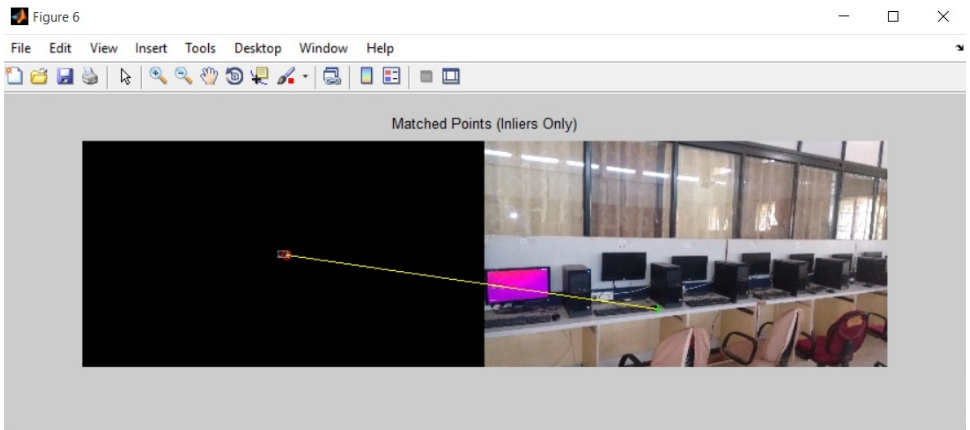
8.3 Image of Object1 (Mouse):





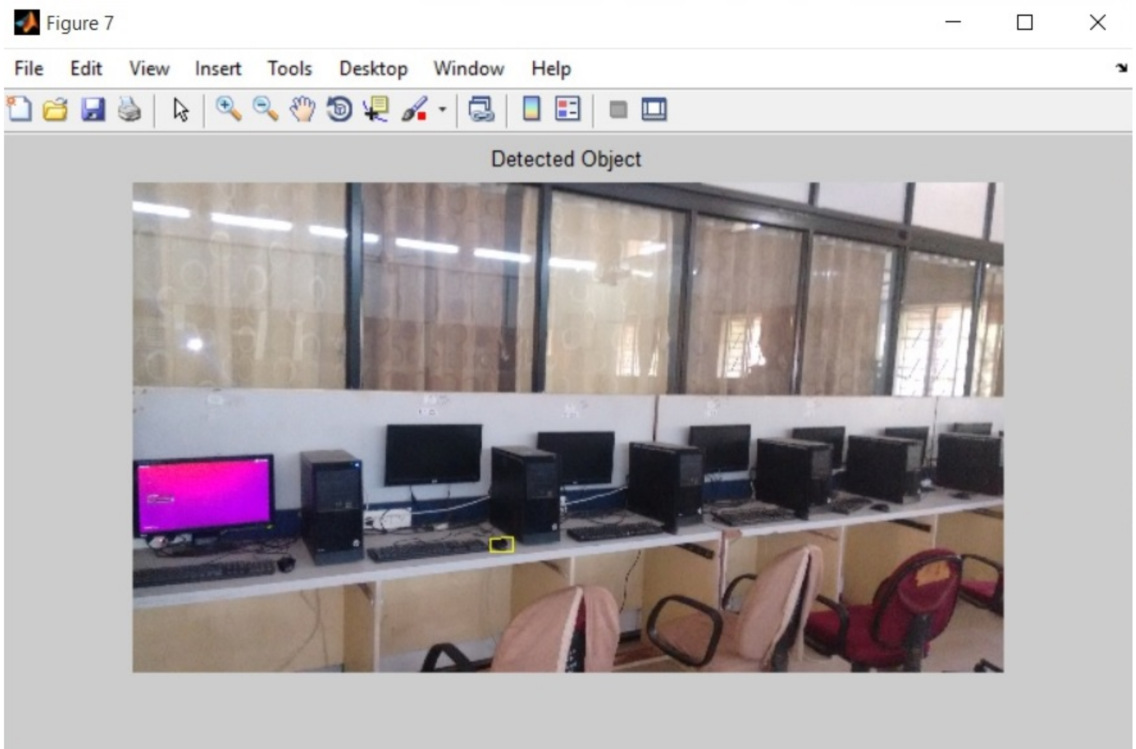
Original image of mouse and the mouse image with strongest feature points are shown in the figures. This is required for mouse object detecting technique.

8.4 Matched features of Mouse:



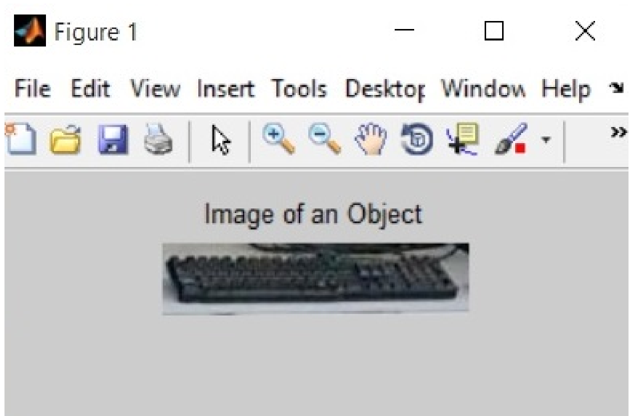
We extracted feature descriptors at the interest points in both the scene and object (mouse) images. Matched feature points using their descriptors from mouse object to the scene image is shown in the figure.

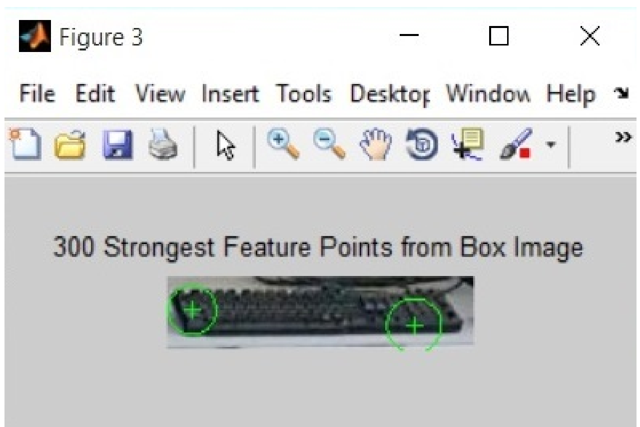
8.5 Detection of Mouse:



After matching the feature points, detected mouse is shown highlighted by a colored boundary box in the scene image.

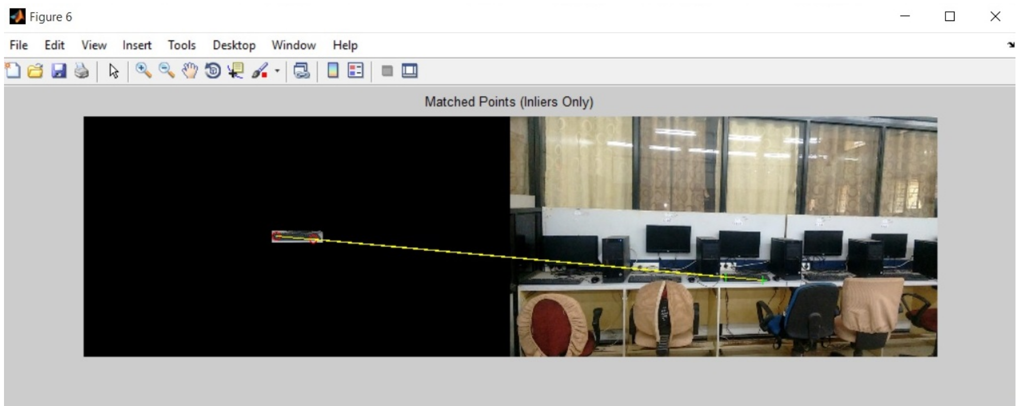
8.6 Image of Object2 (Keyboard):





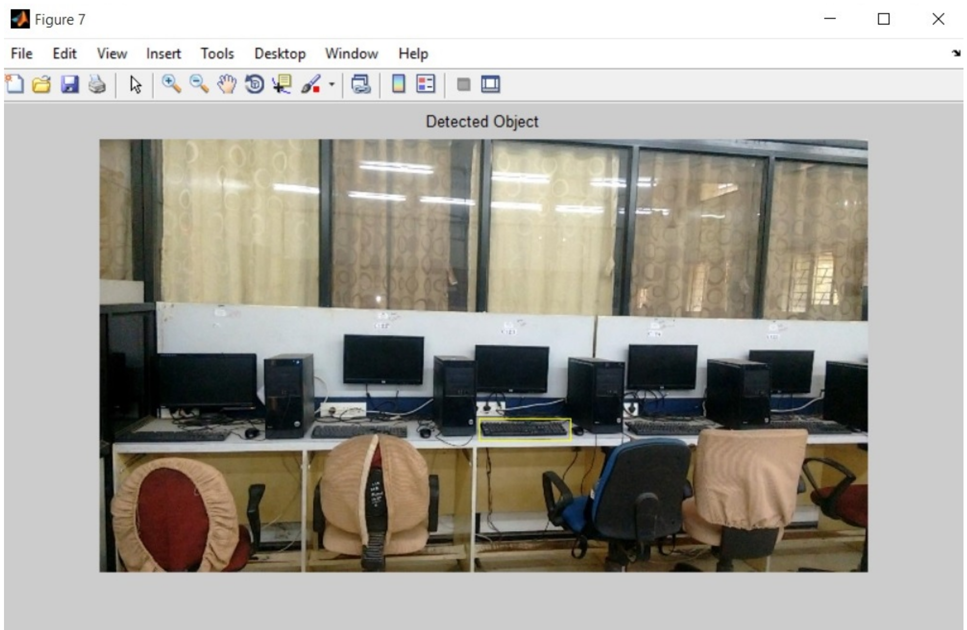
Original image of mouse and the keyboard image with strongest feature points are shown in the figures. This is required for keyboard object detecting technique.

8.7 Matched features of Keyboard :



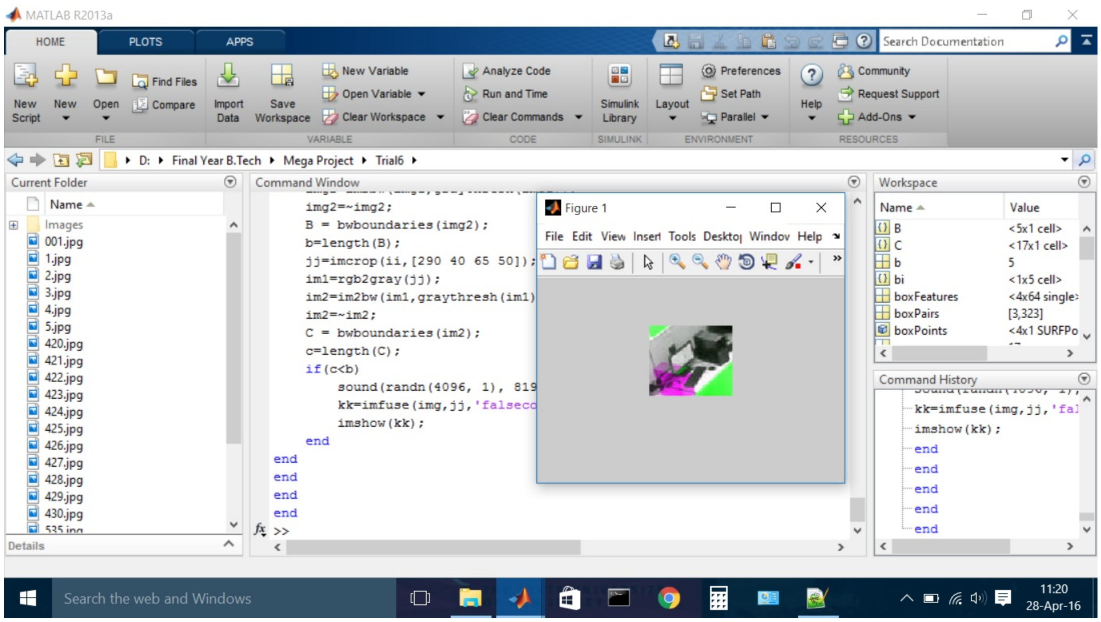
We extracted feature descriptors at the interest points in both the scene and object (keyboard) images. Matched feature points using their descriptors from keyboard object to the scene image is shown in the figure.

8.8 Detection of Keyboard:



After matching the feature points, detected keyboard is shown highlighted by a colored boundary box in the scene image.

8.9 Detection of missing CPU:



Missing CPU in lab is detected and showed in pink color.

9 Conclusion

- The detection of suspicious activities in the department's lab through video surveillance has been successfully implemented for particular lab objects and assets such as CPU, Keyboard, Mouse.
- Integration of object detection and counting objects algorithm has been our major contribution in achieving the objectives of our project. Combination of these two provided a robust solution for detecting missing objects in the lab.
- One of the major drawbacks is the sequential implementation of detection algorithms. Even though it indicates and identifies the missing objects, it does not simultaneously detect the suspicious activities, rather sequentially.
- Poor image resolution was a major challenge we faced while detecting smaller objects in the lab such mouse, external hard disks and pen drives, for which we had to apply image enhancement techniques. Thus, this resulted in an additional step in the overall computation.
- Apart from department's lab, we can implement our project in the Museums, Antique stores and any place holding great significance for objects and materials as this work is mainly object-centric.

10 Future Work

The future work includes implementation of the following:
Further we would extend our project to accomplish the parallelism. The processing time of all the CCTV feeds is huge and complex. Various parallel programming techniques are available with the help of which, the computational time can be reduced. Another important factor is the resolution of current CCTV system. From experimental results, we have concluded that the face detection algorithms could not be implemented on current video frames even after applying various enhancement algorithms as the CCTV suffers from lack of resolution which in turn, fails to recognize authenticated person and passersby as well as small objects such as mouse and keyboard. This issue can be addressed by replacing the current CCTV with other proficient CCTVs present in market. Number of leading CCTV recognition and video analytics products include IndigoVision, HIKVision, Avigilon, and Honeywell. IndigoVision is a leading manufacturer of complete IP video security solutions for airports, ports, rail, traffic, cities, retail, banking, mining, education, casinos, police, prisons and government / local authorities. HIKVision's surveillance products are used in more than 100 countries world-wide. Rapid growth and exceptional products have helped Hikvision become the leading global vendor of CCTV and video surveillance equipment. They have become the leading worldwide supplier of DVR products for three consecutive years. Avigilon's megapixel CCTV camera solutions and their award-winning Avigilon Control Center (ACC) software with High-Definition Stream Management™ (HDSM) technology, deliver pixel-perfect CCTV imagery and unparalleled image detail, leading to faster response and reduced investigation times. Offering a broad range of IP cameras from 1 MP to 29 MP, Avigilon's CCTV cameras come in a variety of formats, including dome, panoramic and fixed. Honeywell cameras provide exceptional imagery. Their video systems allow easy migration from analogue to IP-based technology.

11 References

1. Jaya S. Kulchandan, Kruti J. Dangarwala, “Moving Object Detection: Review of Recent Research Trends”, International Conference on Pervasive Computing (ICPC), IEEE, 2015

2. Akash Gandhamal, Sanjay Talbar, “Evaluation of Background Subtraction Algorithms for Object Extraction”, International Conference on Pervasive Computing (ICPC), IEEE, 2015

3. Abderrahmane EZZAHOUT, Rachid OULAD HAJ THAMI, “Conception and Development of a Video Surveillance System for Detecting, Tracking and Profile Analysis of a Person”, National higher school of Computer Science and Systems Analysis, IEEE, 2013

4. Mohannad Elhamod, Martin D. Levine, “Real-Time Semantics-Based Detection of Suspicious Activities in Public Spaces”, Ninth Conference on Computer and Robot Vision, IEEE, 2012

5. Arnold Wiliem, Vamsi Madasu, Wageeh Boles, Prasad Yarlagadda, “A Context-Based Approach for Detecting Suspicious Behaviours”, 2009 Digital Image Computing: Techniques and Applications

6. Nor Surayahani Suriani, Aini Hussain, Mohd Asyraf Zulkifley, “Sudden Event Recognition: A Survey”, Sensors 2013

7. Dewan Tanvir Ahmed • M. Anwar Hossain •Shervin Shirmohammadi • Abdullah AlGhamdi Pradeep K. Atrey • Abdulmotaleb El Saddik , “Utility based decision support engine for camera view selection in multimedia surveillance systems”, Springer Science+Business Media New York 2012

8. Mukesh Kumar Saini, Pradeep K. Atrey, Abdulmotaleb El Saddik, “From Smart Camera to SmartHub: Embracing Cloud for Video Surveillance”, International Journal of Distributed Sensor Networks, Volume 2014