Smart Surveillance System For Face Recognition

Prepared By

Nivedita Gour

12MCEC05

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

INSTITUTE OF TECHNOLOGY

NIRMA UNIVERSITY

AHMEDABAD-382481

May 2014

Smart Surveillance System For Face Recognition

Major Project

Submitted in partial fulfillment of the requirements

For the degree of

Master of Technology in Computer Science and Engineering

Prepared By

Nivedita Gour

(12MCEC05)

Guided By

Prof. Swati Jain

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

INSTITUTE OF TECHNOLOGY

NIRMA UNIVERSITY

AHMEDABAD-382481

May 2014

Certificate

This is to certify that the Major Project Report entitled "**Smart Surveillance System For Face Recognition**" submitted by **Nivedita Gour (Roll No: 12MCEC05)**, towards the partial fulfillment of the requirements for the degree of Master of Technology in Computer Science and Engineering of Nirma University, 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 Part-II, to the best of my knowledge, haven't been submitted to any other university or institution for award of any degree or diploma.

Prof. Swati Jain
Guide
CSE Department
Institute of Technology
Nirma University, Ahmedabad.

Prof. Vijay Ukani
Associate Professor,
Coordinator M.Tech - CSE,
CSE Department,
Institute of Technology,
Nirma University, Ahmedabad.

Dr. Sanjay Garg
Professor and Head,
CSE Department,
Institute of Technology,
Nirma University, Ahmedabad.

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

Undertaking for Originality of the Work

I, **Nivedita Gour**, Roll. No. **12MCEC05**, give undertaking that the Major Project entitled "**Smart Surveillance System For Face Recognition**" submitted by me, towards the partial fulfillment of the requirements for the degree of Master of Technology in **Computer Science & Engineering** 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 of Student

Date: May , 2014

Place: Ahmedabad

Endorsed by

Prof. Swati Jain

(Signature of Guide)

Acknowledgements

It gives me immense pleasure in expressing thanks and profound gratitude to **Prof. Swati Jain**, Professor, Computer Science Department, Institute of Technology, Nirma University, Ahmedabad for their valuable guidance and continual encouragement throughout this work. The appreciation and continual support they have imparted has been a great motivation to me in reaching a higher goal. Their guidance has triggered and nourished my intellectual maturity that I will benefit from, for a long time to come.

My deepest thanks is extended to **Prof. Vijay Ukani**, PG CSE - Coordinator, Department of Computer Science and Engineering, Institute of Technology, Nirma University, Ahmedabad for an exceptional support and continual encouragement throughout the Major Project.

It gives me an immense pleasure to thank **Dr. Sanjay Garg**, Hon'ble Head of Computer Science and Engineering Department, Institute of Technology, Nirma University, Ahmedabad for his kind support and providing basic infrastructure and healthy research environment.

A special thank you is expressed wholeheartedly to **Dr K Kotecha**, Hon'ble Director, Institute of Technology, Nirma University, Ahmedabad for the unmentionable motivation he has extended throughout the course of this work.

I would also thank the Institution and all faculty members of Computer Engineering Department, Nirma University, Ahmedabad for their special attention and suggestions towards the project work. I would like to thank my colleagues Darpan Korat and Sandeep Kumar for being with me in any condition. I thank them for keeping me cool and calm every time.

- Nivedita Gour

12MCEC05

Abstract

Smart surveillance system refers to video level processing techniques for identification of unwanted(terrorist) faces from real time video. 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)Face Detection, (2)Matching of detected faces with the unwanted faces(terrorist). To detect a face from video frame we use CAMshift algorithm, that gives us subfaces, which can be used by Sift technique for feature extraction and matching with the faces of unwanted person(terrorist).

Further for identifying object as a face from the video of a stationary camera, there are diferent face detection techniques. Once the face detection in video frame is done then the feature extraction and matching to be done.

When face matches with any of unwanted face then the system raise the alarm, so that at sensitive areas like airport, railway station, tourist place etc the security guard or other person get alert tone, thus they can take necessary action and make system secure. Many developed and developing country are using smart surveillance system for viewing the unwanted faces remotely.

Keywords: Smart Surveillance System, Face detection and recognition, CAMshift, Feature Extraction, SIFT technique

Contents

List of Figures

Chapter 1

Introduction

1.1 General

Nowadays, the use of closed-circuit television (CCTV) has increased to secure the premises with the decrease in installation and video storage cost. The excess of terror and crime makes the selective access to place a major concern for many organizations.Conventional methods e.g.password and smart card are unauthentic and fallible. Comparably, face recognition is a reliable and an intelligent biometric identification method.

Automatic face recognition has been a challenging task for the research community. It has been extensively adopted by the applications including biometrics, surveillance, security identification, and authentication. Face recognition usually exploit high-dimensional information which makes it computationally intensive. In addition, wrong detected features can make the recognition process even slower [1].Thus, the interest in robust face recognition techniques to determine whether two facial images belong to same person is increasing rapidly.

Security being popular domain of face recognition, allows us to mount a CCTV camera on fixed position and have a controlled flow of people, thus restricting pose and illumination. Although this reduces the complexity of face recognition, there is still a concern regarding the real time protection of sensitive portion. It should be noted that this problem is somewhat a hard task and can be solved by automatically shooting the unauthorized person attempting to trespass the sensitive area.

Feature observation frameworks have long been being used to screen security touchy areas.the history of feature reconnaissance comprises of three eras of frameworks (era observation frameworks) which are called 1GSS, 2GSS and 3GSS.[2]

1.2 Applications of Surveillance System

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

1.2.1 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.)

1.2.2 Aerial surveillance

Airborne reconnaissance is the social affair of observation, generally visual symbolism or feature, from an airborne vehicle, for example, an unmanned elevated vehicle, helicopter, or spy plane. Military reconnaissance air ship utilize an extent of sensors (e.g. radar) to screen the war zone. Computerized imaging innovation, scaled down workstations, and various other innovative developments over the previous decade have helped fast developments in aeronautical observation equipment, for example, micro-flying vehicles, forward-looking infrared, and high-determination symbolism fit for recognizing articles at amazingly long separations. Case in point, the MQ-9 Reaper, a U.s. automaton plane

utilized for down home operations by the Department of Homeland Security, conveys Polaroids that are fit for distinguishing an item the extent of a milk container from elevations of 60,000 feet, and has forward-looking infrared gadgets that can identify the hotness from a human body at separations of up to 60 kilometers.

1.2.3 Corporate surveillance

Corporate observation is the checking of an individual or gathering's conduct by a corporation.the information gathered is frequently utilized for advertising purposes or sold to other corporations,but is additionally normally imparted to government offices. It might be utilized as a type of business knowledge, which empowers the company to better tailor their items and/or administrations to be alluring by their clients. Then again the information could be sold to different companies, so they can utilize it for the previously stated reason. On the other hand it could be utilized for immediate advertising purposes, for example, the focused on promotions on Google and Yahoo, where ads are focused to the client of the web crawler by examining their inquiry history and messages, which is kept in a database.

1.3 Project Definition

The definition for the project is 'Smart Surveillance System For face recognition'.The main goal of the proposed definition is to develop a prototype consisting of a single camera,for real time detection and identification of terrorist face 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 terrorist face out of them and from that we are detecting the face and then ringing the alarm.

1.3.1 Description

Feature based observation when begun with simply simple CCTV framework obliged human administrators for the transforming of visual data streaming in from frequently various sources.with the movement in innovation and likewise with the huge change in

these frameworks, there still remains the complete reliance on human administrators. As of late, interest for ongoing handling of feature has expanded. For preparing a feature continuously obliges quick and strong calculation. With a specific end goal to meet this necessity it is essential that a reconnaissance framework must be savvy. A definitive objective of Smart Surveillance System (S3) is to permit feature information to be utilized for online alert era to aid human administrators and for offine investigation viably. Such savvy frameworks can produce ongoing cautions characterized on mind boggling occasions and handle dispersed capacity and substance based recovery of feature information. The IBM brilliant reconnaissance framework is one of the few propelled observation frameworks which gives naturally screen a scene as well as the ability to deal with the reconnaissance information, perform occasion based recovery, accept continuous occasion cautions. The IBM S3 is effectively tweaked to the necessity of diverse requisitions. Figure demonstrates the schema of a general shrewd reconnaissance framework.[4]

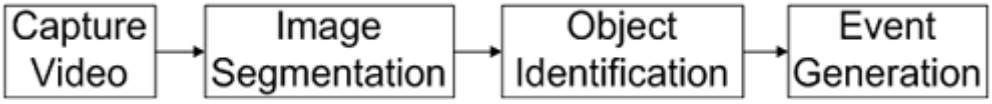

Figure 1.1: Block Diagram of General Smart Video Surveillance System

1.4 Scope of work

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

1.5 Thesis Organization

The work of Smart Surveillance System For Face Recognition and Identification is presented in six chapters.

Chapter two, talks about the existing systems and the literature survey done. It gives

a brief description of the related work and gives an insight to the different methods used for moving object detection, study about face detection and face recognition techniques.
Chapter three, gives the our proposal towards Smart Surveillance System.
Chapter four, illustrates our approach and methodologies used to achieve our proposal. Approach towards face detection using CAMshift and feature extraction using SIFT.
Chapter five, discusses the experimental setup and its result of detecting and matching the faces.
Chapter Six, outlines the conclusion of the work done and future scope which can be carried out further.

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 utilization of programmed picture understanding method to concentrate data from the observation information.

Pictures are recognized as a standout amongst the most essential medium of passing on data. A vital part of machine taking in is to comprehend the picture and to concentrate data out of it. To comprehend a picture the first step is to portion it and discover distinctive questions in it.although division is recognized to be one of the essential steps in article detection,it is additionally acknowledged to be a standout amongst the most prominent issues in workstation vision. Byoung in his audit of past related studies, ordered these procedures into taking after: thresholding methodologies, shape based methodologies, district based methodologies, grouping based methodologies and other streamlining based methodologies utilizing a Bayesian framework,neural systems. A couple of division systems for article recognition have been examined underneath.[5].

2.1 Video Surveillance System

Min Moon, Chulho Won, Sung Bum Pan in their paper "Multi-Modal Human Identication 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.[6]

Thiago T. Santos , Carlos H. Morimoto in their paper "Numerous Polaroid individuals discovery and following utilizing help incorporation" have proposed a technique to spot and track individuals by consolidating proof from different Polaroids utilizing homography imperative. The proposed strategy utilization frontal area pixels from basic foundation subtraction to figure proof of the area of individuals on a reference ground plane. The calculation processes the measure of help that essentially compares to the "frontal area mass" over every pixel. In this manner, pixels that compare to ground focuses have more backing. The help is standardized to make up for point of view impacts and amassed on the reference plane for all Polaroid sees. The discovery of individuals on the reference plane turns into a quest for locales of nearby maxima in the ccumulator. A lot of people false positives are separated by checking the perceivability consistency of the located competitors against all Polaroid sees. The remaining hopefuls are followed utilizing Kalman channels and appearance models. The principle commitment of this paper is the meaning of a novel calculation focused around the homography demand that does not depend on single perspective division of the subjects or past following data.[**?**]

Chung-Hao Chen a, Yi Yao b, David Page c, Besma Abidi d, Andreas Koschan d, Mongi Abidi[7] in their paper "Polaroid handoff with versatile asset administration for multi-Polaroid multi-item following" have outlined a handoff calculation to acquire a persistently followed and reliably marked trajectory of the object of enthusiasm toward multi-Polaroid observation framework. Polaroid handoff is a choice methodology of exchanging a versatile item starting with one Polaroid then onto the next, wherein predictable marking tackling the character issue among numerous watching Polaroids and establishing the framework for Polaroid handoff. All in all, Polaroid handoff controls the coordinated effort among different Polaroids and responses the inquiries of When and Who: when a handoff appeal ought to be activated to secure sufficient time for a fruitful

reliable marking and who is the most qualified Polaroid to assume control over the object of enthusiasm before it drops out of FOV of at present watching Polaroid. Most existing handoff calculations pick the Polaroid to which the object of investment is approaching. This straightforward standard is often deficient and prompts 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 [7] in their paper "Color picture division utilizing an upgraded Gradient Network Method" have assessed another methodology for solid shade picture division, in pictures giving color structure solid yet persistent shade changes. Joins an improved adaptation of Gradient Network 2, with basic district developing strategy utilized as presegmentation.

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 [**?**] 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 [**?**] , in their paper "Feature Object Segmentation Based on Accumulative Frame Difference" have tended to the issue of concentrating feature objects from head-shoulder feature groupings. They have proposed a strategy focused around Accumulative Frame Difference (AFD). At first by performing movement investigation on each one piece of frame,the squares with quick moving edges are recognized. At that point, for each one square, they have amassed the casing distinction with an alternate measure of edges, taking into account its movement traits. In the wake of thresholding and post transforming, the items are gotten.

Yasira Beevi C P, Dr. S. Natarajan[8], in their paper "A proficient Video Segmentation Algorithm with Real time Adaptive Threshold Technique" have displayed a feature division calculation for MPEG-4 Polaroid framework with change discovery, foundation enlistment strategies and constant versatile limit procedures. Their work likewise comprise of a shadow scratch-off mode which can manage light changing impacts and shadow impacts.

2.3 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.

2.3.1 K-means Clustering Algorithm

The bunching methodologies could be classified into two general gatherings: partitional and various leveled grouping calculations. Partitional bunching calculations, for example, K-means and EM grouping are generally utilized within numerous requisitions, for example, information mining, packing, picture division, and machine taking in. Along these lines, the focal point of grouping calculations is that the characterization is straightforward and simple to actualize. Likewise, the burdens are of how to focus the amount of groups and diminishing the amounts of emphasis.

The K-Means is a nonhierarchical grouping method that takes after a straightforward strategy to order a given information set through a specific number of K bunches that are known from the earlier. The K-Means calculation redesigns the space part of the info information iteratively, where the components of the information are traded between bunches focused around a predefined metric (ordinarily the Euclidian separation between the group focuses and the vector under examination) so as to fulfill the criteria of minimizing the variety inside each one bunch and amplifying the variety between the ensuing K groups.

Steps of the traditional K-Means bunching calculation:

- **Initialization** - produce the beginning condition by characterizing the amount of bunches and haphazardly select the introductory group focuses.
- Generate another segment by allocating every information point to the closest bunch focus.
- Recalculate the communities for groups getting new information focuses and for cluste rs losing information focuses.
- Repeat the steps 2 and 3 until a separation meeting foundation is met.

2.3.2 Region Growing Algorithm

The primary locale developing system was the seeded district developing technique. This technique takes a set of seeds as info alongside the picture. The seeds check each of the items to be sectioned. The areas are iteratively developed by contrasting all unallocated neighboring pixels with the locales. The distinction between a pixel's force worth and the

locale's mean is utilized as a measure of similitude. The pixel with the littlest distinction measured thusly is allotted to the individual area. This methodology proceeds until all pixels are dispensed to a locale. Seeded district developing obliges seeds as extra include.

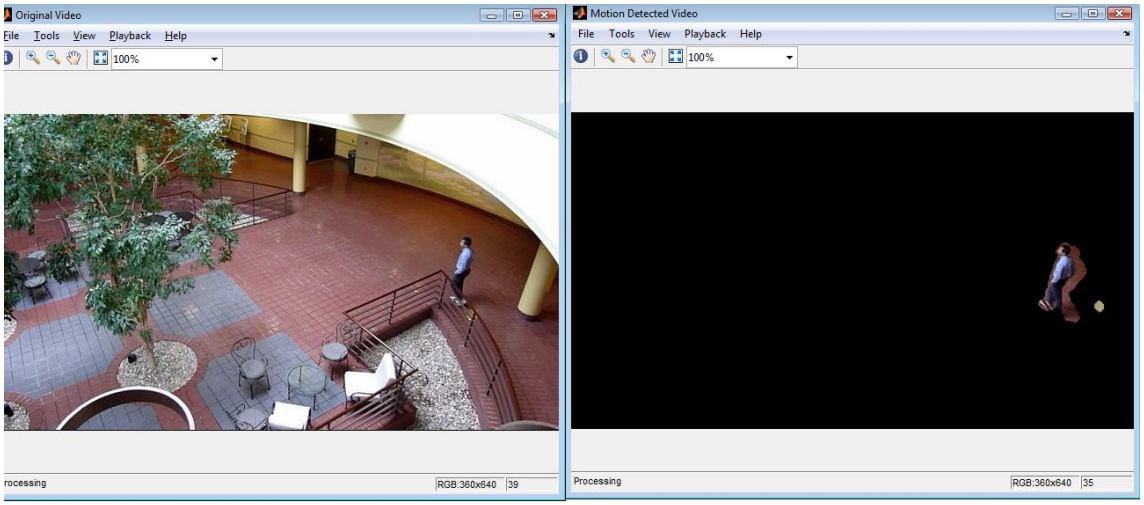

Figure 2.1: video frame and their results after applying region growing algorithm

Figure 2.2: another video frames and their results after applying region growing algorithm

The division results are reliant on the decision of seeds. Clamor in the picture can result in the seeds to be crudely put. Unseeded area developing is an altered calculation that doesn't require express seeds. Figure indicates the results acquired on applying the area developing calculation. The results delineate that the area developing calculation does not prepare proper brings about constant. Locale developing division systems display two principle offers that point of confinement their appropriateness for managing proficiently with characteristic scenes.[**?**, e] A static district likeness idea: where pixels or surfaces inside an area are relied upon to be homogeneous. Commonplace characteristic scenes, be that as it may, demonstrate solid ceaseless varieties of shade, exhibiting an

alternate, element request that is not considered by such calculations.

Build in many-sided quality to present more steady comes about: which normally requests complex processings to locate section correspondence signs, or are based upon extra surface data. This backs off significantly the preparing time without being substantially more steady when compelling color varieties are available.

2.3.3 Background Subtraction Method

A straightforward and basic movement recognition technique includes subtracting every new picture It(x;y) in a feature succession from a model of the foundation scene void of articles Bt(x;y) and thresholding the ensuing contrast, highlighting the frontal area pixels.the pixels having the qualities above edge are named forefront pixels. Thus a paired picture is structured.[3].

Figure 2.3: (a)Background image, (b)different frames of a video

$$|I_t(x,y) - B_t(x,y)| > T$$

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

Figure 2.4: Result of background subtraction

Figure2.4 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.Figure2.5 shows six different frames of a video sequence.[**?**]

Figure 2.5: Background frame of the video sequence

Figure 2.6: Frames containing some moving object

2.4 Face Detection : Different Methods

The face detection means to determine the position of the face in the image, the size, position and attitude of the process. It is a development direction and important application of pattern recognition, as a key technology in the information processing of human faces, and has been active research in the field of pattern recognition and computer vision issues.

FOUR TYPES OF FACE DETECTION METHODS AND SEVERAL CORRESPONDING ALGORITHMS CONTRAST

2.4.1 Knowledge-based face detection

Knowledge-based Face Detection technology face denomination organs relationship between coding guidelines of face detection technology, which is the top-down, according

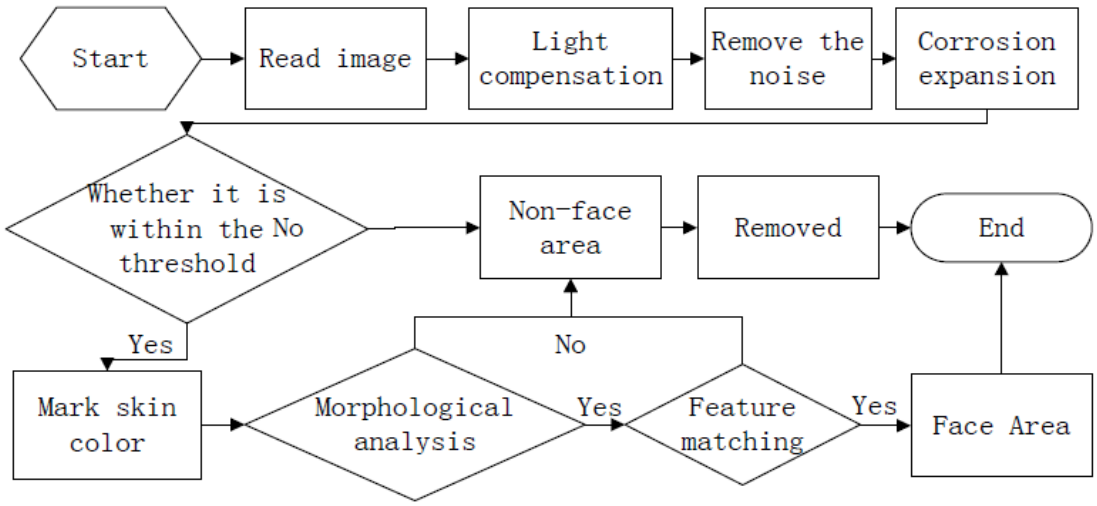

Figure 2.7: The Process of Face Detection

to the symmetry of the human facial organs, gray-scale differences in prior knowledge to develop a series of criteria. When the image in the test area meet the criteria, it is detected human face. 1994 Yang et al.[3] proposed mosaic-based face detection technology, they use a 44 mosaic of the face block, to develop criteria to determine the gray value of each block. The system is divided into three, using the average and the second sampling of different precision three different resolution images. Using different criteria to determine different resolution images. In the low-resolution images, the guidelines reflect the broad contours of the face. If the high-resolution image, ready to reflect the face of the minutiae. In 1999, Chunyu Lu et al.[4] improved the mosaic technology, according to the natural distribution of the face organs they divided face into 33 mosaic block, adaptively adjusted the size of the blocks in the detection, test to be detected region using a set of guidelines based on knowledge of each block gray and gradient statistical characteristics whether it is someone's face. Jun Jiang et al.[5] proposed a new generalized geometric projection method, on the basis of the face of victory structural features and Chunyu Lu's 33three-point diagram They established a relatively complete knowledge base of the face, comprehensive grayscale rules, gradient rules and binary Rules of,which broadens the range of use of the detection algorithm.

2.4.2 Feature-based face detection

Feature-based method can detect the face not only from the existing facial features but also from their geometric relationship. It is contrary to the knowledge-based method.

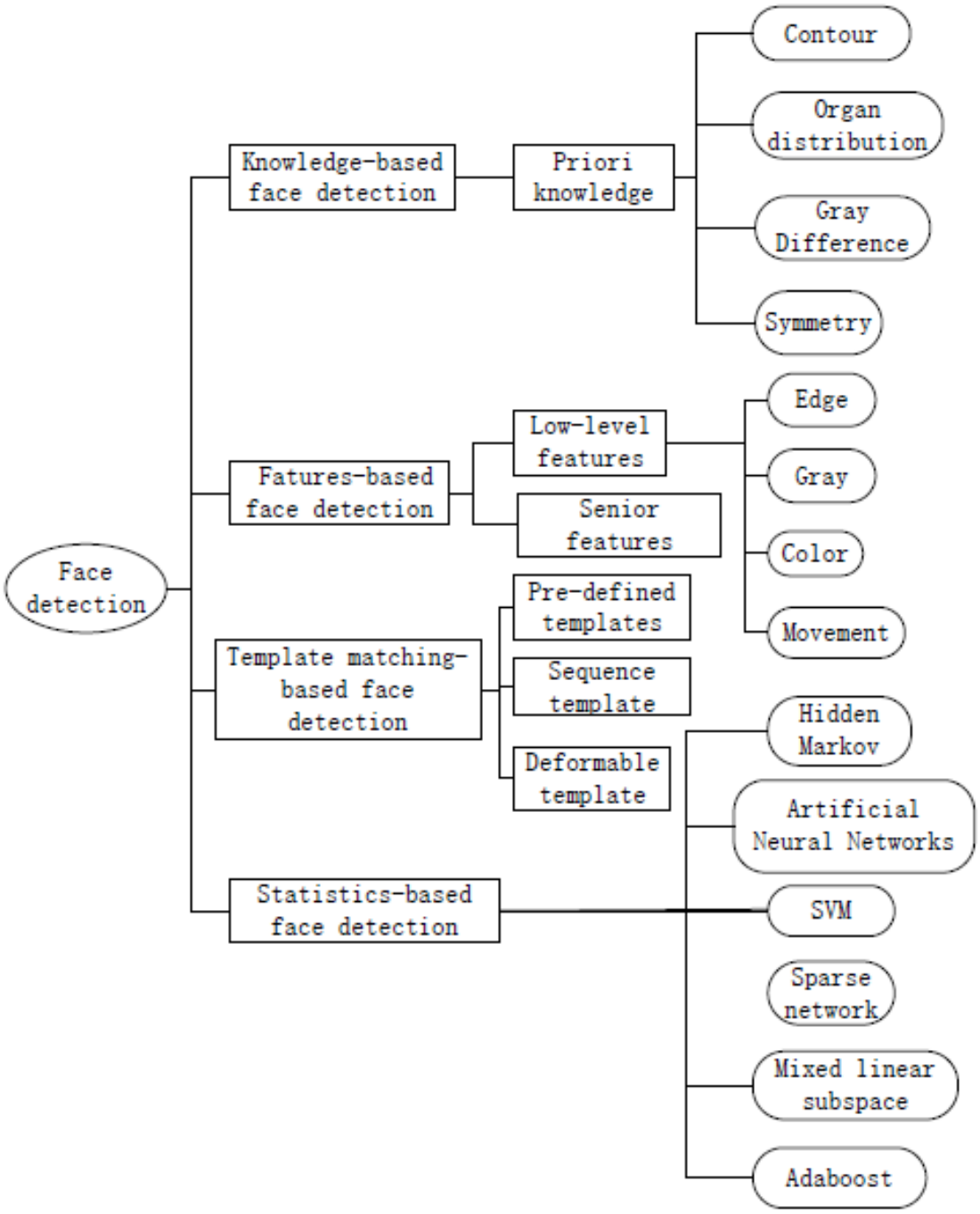

Figure 2.8: Division of Different Methods

It is the first to use a variety of means to find invariant features of the face, and then integrated invariant feature to determined the region to be detected whether it is a human face. Invariant features of the face: eyebrows, eyes, nose, mouth and hairline. We usually use the edge detector to extract them, and then create a statistical model to describe the relationship between the characteristics based on the extracted features, and to determine the presence of human face.

2.4.3 Template-based face detection

Template-based method can be divided into two categories: predetermined template and deformable template. Predetermined template method develops a standard template, and then calculate the associated value of the detection area and templates, when the associated value is in line with the guidelines established by, the detection area is human face. Deformable template first to develop a template parameter, and then, according to the detection area to modify parameters until convergence, in order to achieve the purpose of the location of the detected facial organ.

2.4.4 Statistics-based face detection

Such methods takes face region as a class model, namely the template feature, using a lot of "face" and "non-face training samples, constructed classifier to distinguish all possible areas in the images belong to the types of mode face detection. Currently it is widely used in statistical methods and face detection are: subspace-based methods, neural networks, support vector machine(SVM), Hidden Markov and AdaBoost algorithm.

2.5 Feature Extraction : Approaches

Characteristic extraction includes rearranging the measure of assets needed to portray an extensive set of information correctly. At the point when performing dissection of complex information one of the significant issues comes from the amount of variables included. Dissection with an expansive number of variables by and large obliges a lot of memory and processing force or an order calculation which overfits the preparation test and sums up inadequately to new examples. Characteristic extraction is a general term for strategies for building blends of the variables to get around these issues while even now portraying the information with sufficient precision.

2.5.1 Principal Component Analysis

A standout amongst the most utilized and refered to factual technique is the Principal Component Analysis (PCA) [?] [?] [?]. It is a numerical strategy that performs a dimensionality diminishment by concentrating the key parts of the multi-dimensional information. The principal vital part is the direct mix of the first measurements that has the most astounding variability. The n-th key segment is the direct mix with the most extreme variability, being orthogonal to the n-1 first chief segments.

Typically the mean $\bar{x}$is separated from the information. So,let Xn Xm be the information grid where X1, ..., Xm are the picture vectors (vector sections) and n is the amount of pixels for every picture. The KLT premise is acquired by tackling the eigenvalue issue.

$$C_x = \phi \Lambda \phi^T \tag{2.1}$$

where Cx is the covariance matrix of the data

$$C_x = \frac{1}{m} \sum_{i=1}^{m} x_i x_i^T \tag{2.2}$$

Φ=[ϕ_1,.....ϕ_n] is the eigenvector matrix of Cx. $\Lambda$is a diagonal matrix, the eigenvalues λ_1, ...,λ_n n of Cx are located on its main diagonal. λi is the variance of the data projected on ϕi.

2.5.2 Local Binary Pattern

The first LBP administrator, presented by Ojala et al [?], is a compelling method for surface portrayal. The administrator names the pixels of a picture by thresholding the 3x3-area of every pixel with the middle esteem and recognizing the result as a double number. At that point the histogram of the marks could be utilized as a composition descriptor. See Figure 3.2 for a delineation of the essential LBP administrator. Later the administrator was stretched out to utilize neigbourhoods of distinctive sizes [10]. Utilizing roundabout neighborhoods and bilinearly introducing the pixel qualities permit any sweep and number of pixels in the area. For neighborhoods we will utilize the documentation (P, R) which implies P inspecting focuses on a round of span of R. See Figure 3.3 for a sample of the roundabout (8,2) area. An alternate expansion to the first administrator

utilizes supposed uniform examples. [10].A Local Binary Pattern is called uniform in the event that it holds at most two bitwise moves from 0 to 1 or the other way around when the double string is viewed as round. Case in point, 00000000, 00011110 and 10000011 are uniform examples. Ojala et al. Perceived that in their explores different avenues regarding composition pictures, uniform examples represent a bit short of what 90% of all examples when utilizing the (8,1) area and for around 70% in the (16,2) area.

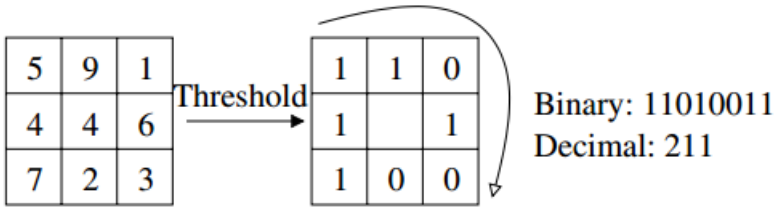

Figure 2.9: The basic LBP operator.

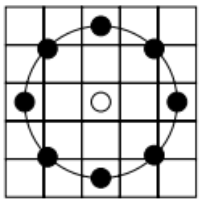

Figure 2.10: The roundabout (8,2) neigbourhood. The pixel qualities are bilinearly interjected at whatever point the inspecting point is not in the core of a pixel.

We utilize documentation for the LBP administrator $lbp^u2_{p,r}$ The subscript speaks to utilizing the administrator as a part of a (P, R) neighborhood. Superscript u2 remains for utilizing just uniform examples and naming all remaining examples with a solitary name. A histogram of the marked picture fl(x, y) could be characterized as:

$$H_i = \sum_{x,y} \tag{2.3}$$

I fl(x, y) = i, i = 0,...,n-1,

in which n is the amount of diverse marks transformed by the LBP administrator and Ia=if 1 An is TRUE else for 0 it is FALSE. This histogram holds data about the dispersion of the nearby micropatterns, for example, edges, spots and level ranges, over the entire picture. For proficient face representation, one ought to hold likewise spatial

data. For this reason, the picture is separated into districts R0,r1, . . . Rm-1 and the spatially improved histogram is characterized as:

$$H_i = \sum_{x,y} \tag{2.4}$$

I fl(x, y) = i I (x, y) ϵ Rj , i = 0, . . . , n.1, j = 0, . . . , m-1. In this histogram, we adequately have a depiction of the face on three separate levels of area: the names for the histogram hold data about the examples on a pixel-level, the marks are summed over a little district to generate data on a territorial level and the local histograms are connected to assemble a worldwide portrayal of the face.

2.5.3 Linear Discriminant Analysis

LDA is generally used to discover direct mixes of characteristics while protecting class distinguishableness. Dissimilar to PCA, LDA tries to model the contrasts between classes. Fantastic LDA is intended to consider just two classes. Particularly, it obliges information indicates for distinctive classes be a long way from one another, while point from the same class are close. Subsequently, LDA acquires differenced projection vectors for each one class. Multi-class LDA calculations which can oversee more than two classes are more utilized.

2.6 Face recognition : Different approaches

Face distinguishment is an advancing region, changing and enhancing always. Numerous exploration regions influence face distinguishment - workstation vision, optics, design distinguishment, neural systems, machine taking in, psycology, and so on. Past segments illustrate the diverse steps of a face distinguishment process. There is not an agreement on that respect. All these components block the improvement of a brought together face distinguishment calculation arrangement plan.

2.6.1 Geometric- Template Based approaches

Face distinguishment calculations could be named either geometry based or layout based calculations [**?**] [**?**]. The layout based techniques contrast the information picture and a set of layouts. The set of layouts might be costructedusing measurable apparatuses like Suppor Vector Machines (SVM) [**?**], Principal Component Analysis (PCA) [**?**] [**?**] [**?**],

Linear Discriminant Ananlysis (Lda)[?], Independent Component Analysis (ICA) [?] [?] [?], Kernel Methods, or Trace Transforms. The geometry characteristic based systems examine nearby facial characteristics and their geometric connections. This methodology is frequently called characteristic based methodology [12]. There are calculations created utilizing both methodologies. Case in point, a 3d morphable model methodology can utilize characteristic focuses or composition and in addition PCA to construct a distinguishment framework. [?]

2.6.2 Appeareance-based or Model-based approaches

Facial recoginition strategies could be isolated into appeareance-based or model based calculations. The differential component of these techniques is the representation of the face. Appeareance-based strategies speak to a face as far as a few crude force pictures. A picture is recognized as a high-dimensional vector. At that point satistical techiques are generally used to infer a characteristic space from the picture appropriation. The example picture is contrasted with the preparation set. Then again, the model-based methodology tries to model a human face. The new specimen is fitted to the model, and the parametres of the fitten model used to perceive the picture.

Straight appeareance-based techniques perform a direct measurement decrease. The face vectors are anticipated to the premise vectors, the projection coefficients are utilized as the characteristic representation of each one face picture. Illustrations of this methodology are PCA, LDA or ICA. Non-straight appeareance routines are more muddle. Actually, direct subspace investigation is a close estimation of a nonlinear complex.

2.6.3 Statistical approach for recognition algorithms

Pictures of confronts, spoke to as high-dimensional pixel shows, regularly have a place with a complex of easier measurement. In measurable methodology, each one picture is spoken to regarding d characteristics. Along these lines, it's saw as a point (vector) in a d-dimensional space. The dimensionality number of directions required to determine an information purpose of this information is excessively high. Along these lines, the objective is to pick and apply the right factual device for extraction and investigation of the underlaying complex. These apparatuses must characterize the implanted face space in the picture space and concentrate the premise capacities from the face space.

This might allow examples having a place with distinctive classes to involve disjoint and compacted areas in the characteristic space. Hence, we might have the capacity to characterize a line, bend, plane or hyperplane that differentiates confronts fitting in with diverse classe

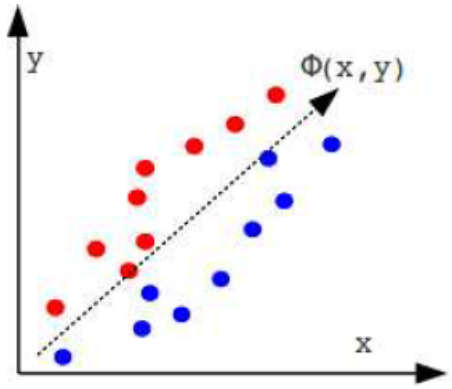

Figure 2.11: PCA. x and y are the first basis.ϕ is the first essential part.

Chapter 3

Proposal

Currently, in Smart Surveillance System, the stationary cameras are just recording a video, and store it in their database. These stored videos are later on viewed if necessary or to be checked that who has done any malfunction. Thus we later on come to know who has done such malfunction activity. Therefore we need a Surveillance system in which on detecting any of such person's face, it should give us a signal or raise an alarm.

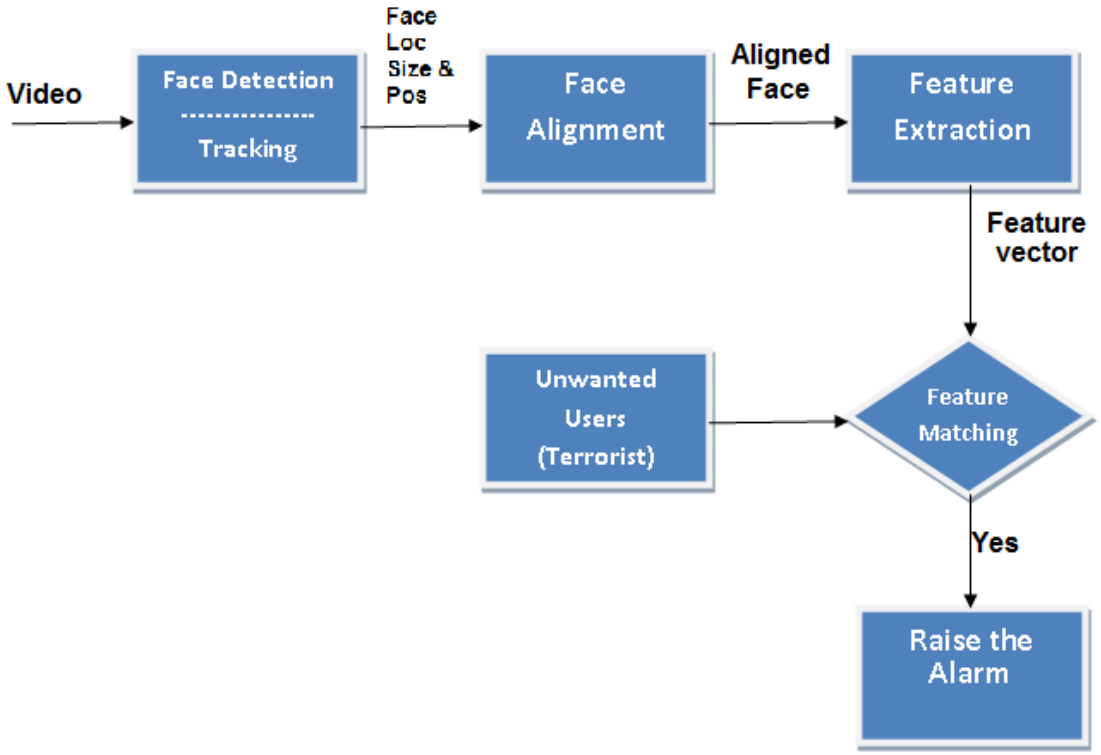

Figure 3.1: Data Flow of the Proposal

In order to develop a Smart Surveillance System, first we consider that a stationary camara is capturing video and that video is further processed. This process takes video as input and generate the different different frames. Each frame may contain multiple faces. Each face's location, size and position is different, therefore the face alignment process is to be done. After face alignment the subfaces are matched by a feature extraction or matching technique with a already build dataset of unwanted users or Terrorist. If any of the face is matched with the unwanted user or Terrorist then alarm is raised. Only terrorist faces to be matched, in case of other person, no alarm should raise and further process with new frames to be processed.

Chapter 4

Approaches and Methodology

4.1 Camshift

Camshift is called Continuously Adaptive Mean Shift focused around the mean movement calculation [2]. Camshift utilizes the Hue channel to follow objects subsequent to by utilizing the Hue channel focused around HSV shade model, objects with distinctive colors might be perceived. In view of the color data, Camshift tracks questions speedier and devours generally little CPU assets. More level registering asset necessity empower Camshift to turn into an one of constant face following calculations.

4.1.1 proposed algorithm for CAMshift

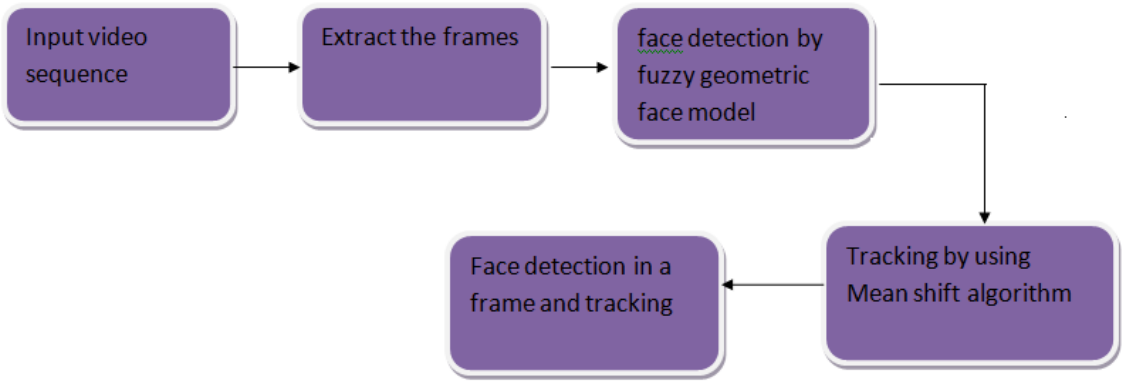

Figure 4.1: Block diagram of the proposed approch

The strategy of Camshift demonstrated in the Figure, incorporates two vital parts which are the histogram and the hunt of top likelihood. The primary venture of Camshift is to get the histogram of followed article. In the second step, the following casing will be

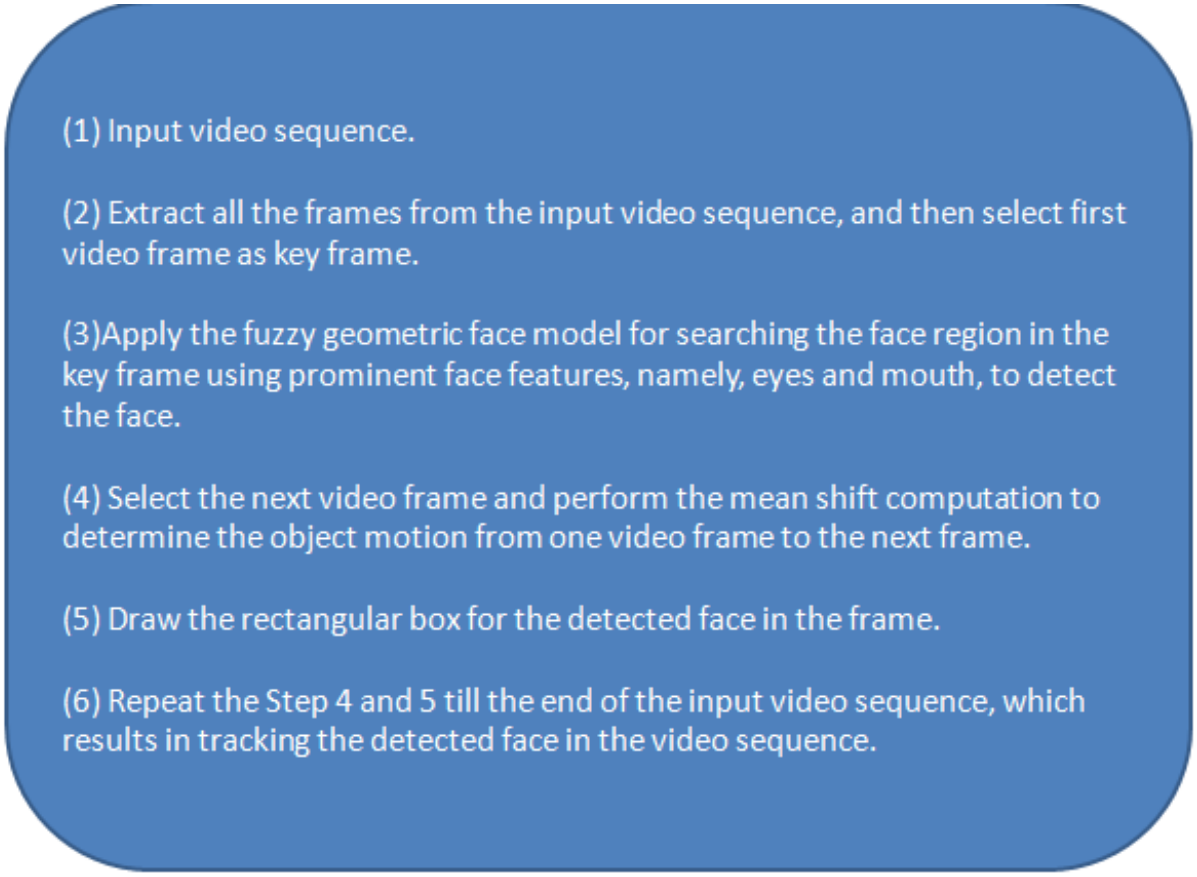

Figure 4.2: proposed algorithm for CAMShift

changed over into a guide of skin color likelihood focused around the histogram. In the third step, the top likelihood focus will be discovered focused around zero minute and first minute. At long last, Camshift will check whether it focalizes. In the event that it is no, it will go to the third step; else, it gets the position of the followed question in this casing and brings the following edge to track the item ceaselessly.

4.2 Face detection : using CAMshift

Object identification and following are paramount in numerous workstation vision provisions including action distinguishment, auto security, and observation. In this case, you will create a straightforward face following framework by separating the following issue into three different issues:

- Detect a face to track
- Identify facial characteristics to track
- Track the face

4.2.1 Detect a Face To Track

Before you start following a face, you have to first identify it. Utilize the vision. Cascadeobjectdetector to catch the area of a face in a feature outline. The course question locator utilizes the Viola-Jones location calculation and a prepared order model for discovery. Naturally, the locator is designed to catch faces, however it might be arranged for other article sorts.

You can utilize the course protest identifier to track a face crosswise over progressive feature outlines. Be that as it may, when the face tilts or the individual turns their head, you may lose following. This limit is because of the sort of prepared order model utilized for location. To evade this issue, and on the grounds that performing face identification for each feature edge is computationally concentrated, this illustration utilizes a straightforward facial characteristic for following.

4.2.2 Identify Facial Features To Track

When the face is found in the feature, the following step is to distinguish a characteristic that will help you track the face. For instance, you can utilize the shape, surface, or color. Pick a characteristic that is remarkable to the article and stays invariant actually when the item moves.

In this illustration, you utilize skin tone as the characteristic to track. The skin tone gives a great arrangement of complexity between the face and the foundation and does not change as the face turns or moves.

4.2.3 Track the Face

With the skin tone chose as the characteristic to track, you can now utilize the vision.histogrambasedtracker for following. The histogram based tracker utilizes the Camshift calculation, which gives the competence to track an article utilizing a histogram of pixel qualities.

In this sample, the Hue channel pixels are concentrated from the nose locale of the identified face. These pixels are utilized to instate the histogram for the tracker. The case tracks the article over progressive feature casings utilizing this histogram.

4.3 SIFT : Scale Invariant Feature Transforms

Image matching is an essential part of numerous issues in machine vision, including item or scene distinguishment, explaining for 3d structure from various pictures, stereo correspondence, and movement following. This paper portrays picture emphasizes that have numerous properties that make them suitable for matching contrasting pictures of an item or scene. The characteristics are invariant to picture scaling and revolution, and part of the way invariant to change in brightening and 3d Polaroid perspective. They are overall confined in both the spatial and recurrence spaces, diminishing the likelihood of disturbance by impediment, mess, or clamor. Substantial amounts of characteristics might be concentrated from regular pictures with effective calculations.

Also, the characteristics are exceptionally different, which permits a solitary characteristic to be rightly matched with high likelihood against an expansive database of characteristics, giving a premise to question and scene distinguishment. The expense of concentrating these characteristics is minimized by taking a course sifting methodology, in which the more unreasonable operations are connected just at areas that pass a starting test. Taking after are the real phases of calculation used to produce the set of picture characteristics:

- **Scale-space extrema detection :** The primary phase of calculation ventures over all scales and picture areas. It is executed proficiently by utilizing a distinction of-Gaussian capacity to recognize potential investment indicates that are invariant scale and introduction.

  This phase of the separating endeavors to recognize those areas and scales that are identifiable from distinctive perspectives of the same article. This might be effectively attained utilizing a "scale space" capacity. Further it has been indicated under sensible suspicions it must be focused around the Gaussian capacity. The scale space is characterized by the capacity:

$$L(x, y, \sigma) = G(x, y, \sigma) * I(x, y)$$

  Where * is the convolution administrator, $g(x, y, \sigma)$ is a variable-scale Gaussian and $i(x, y)$ is the info picture.

Different systems can then be utilized to distinguish stable keypoint areas in the scale-space. Contrast of Gaussians is one such system, finding scale-space extrema, $D(x, y, \sigma)$ by figuring the distinction between two pictures, unified with scale k times the other. $D(x, y, \sigma)$ is then given by:

$$D(x, y, \sigma) = L(x, y, k\sigma) - L(x, y, \sigma)$$

To catch the nearby maxima and minima of $d(x, y, \sigma)$ each one point is contrasted and its 8 neighbors at the same scale, and its 9 neighbors all over one scale. In the event that this worth is the base or greatest of all these focuses then this point is an extrema.

- **Keypoint localization :** At every hopeful area, a nitty gritty model is fit to focus area and scale. Keypoints are chosen focused around measures of their security.

  This stage endeavors to dispense with more focuses from the arrangement of keypoints by discovering those that have low complexity or are inadequately restricted on an edge. This is accomplished by figuring the Laplacian, esteem for every keypoint found in stage 1. The area of extremum, z, is given by:

$$Z = \frac{\partial^2 D^{-1}}{\partial x^2} \frac{\partial D}{\partial x}$$

  On the off chance that the capacity esteem at z is underneath a limit esteem then this point is prohibited. This evacuates extrema with low complexity. To take out extrema focused around poor localisation it is noted that in these cases there is a huge standard curve over the edge yet a little arch in the perpendicular course in the defference of Gaussian capacity. On the off chance that this distinction is beneath the degree of biggest to most modest eigenvector, from the 2x2 Hessian network at the area and scale of the keypoint, the keypoint is rejected.

- **Orientation assignment :** One or more introductions are relegated to every keypoint area focused around nearby picture slope headings. All future operations are performed on picture information that has been converted with respect to the appointed introduction, scale, and area for each one characteristic, along these lines giving invariance to these conversions.

This step expects to relegate a reliable introduction to the keypoints focused around neighborhood picture properties. The keypoint descriptor, depicted beneath, can then be spoken to with respect to this introduction, accomplishing invariance to revolution. The methodology taken to discover an introduction is:

- Utilize the keypoints scale to select the Gaussian smoothed picture L, from above
- Compute angle magnitude,m

$$m(x,y) = \sqrt{(L(x+1,y) - L(x-1,y))^2 + (L(x,y+1) - L(x,y-1))^2}$$

- Compute orientation, θ

$$\theta(x,y) = tan^2((L(x,y+1) - L(x,y-1))/(L(x+1,y) - L(x-1,y)))$$

- Form an introduction histogram from angle introductions of specimen focuses
- Locate the most elevated top in the histogram. Utilize this crest and whatever viable neighborhood top inside 80% of the stature of this top to make a keypoint with that introduction
- Some focuses will be relegated numerous introductions
- Fit a parabola to the 3 histogram values closest to each one crest to add the tops position.

- **Keypoint descriptor :**

The nearby picture angles are measured at the chose scale in the area around every keypoint. These are changed into a representation that considers noteworthy levels of nearby shape bending and change in enlightenment.

The nearby angle information, utilized above, is likewise used to make keypoint descriptors. The slope data is turned to line up with the introduction of the keypoint and afterward weighted by a Gaussian with fluctuation of 1.5 * keypoint scale. This information is then used to make a set of histograms over a window fixated on the keypoint.

Keypoint descriptors regularly utilizes a set of 16 histograms, adjusted in a 4x4 lattice, each with 8 introduction containers, one for each of the principle compass

bearings and one for each of the mid-purposes of these headings. This results in a characteristic vector holding 128 components.

These ensuing vectors are know as SIFT keys and are utilized as a part of a closest neigbours methodology to recognize conceivable protests in a picture. Accumulations of keys that concede to a conceivable model are identfied, when 3 or more keys coincide on the model parameters this model is apparent in the picture with high likelihood. Because of the expansive number of SIFT keys in a picture of an article, regularly a 500x500 pixel picture will create in the locale of 2000 characteristics, significant levels of impediment are conceivable while the picture is still perceived by this procedure.

Chapter 5

Experimental Setup and Results

5.1 Implementation Methodology

NOTE: In the work presented here, Matlab 12 has been used for implementing the algorithm. It is considered that the camera is stationary. Videos have been recorded using a single camera with the resolution of 640x480. The system running this algorithm is fast enough to process heavy load of algorithm and dataset of unwanted people is already built. The alarm system has also enough power to make all security guard alert at some distence.

5.1.1 Face detection

When the video input is given to the system. The system checks the type of video and the frame rate and the frame size. Each frame is considered as the photo frame and by using CAMshift algorithm it is checked for any face present in that frame. If any face is detected, then the localization of face is done in the form of ractangle, as shown in figure:

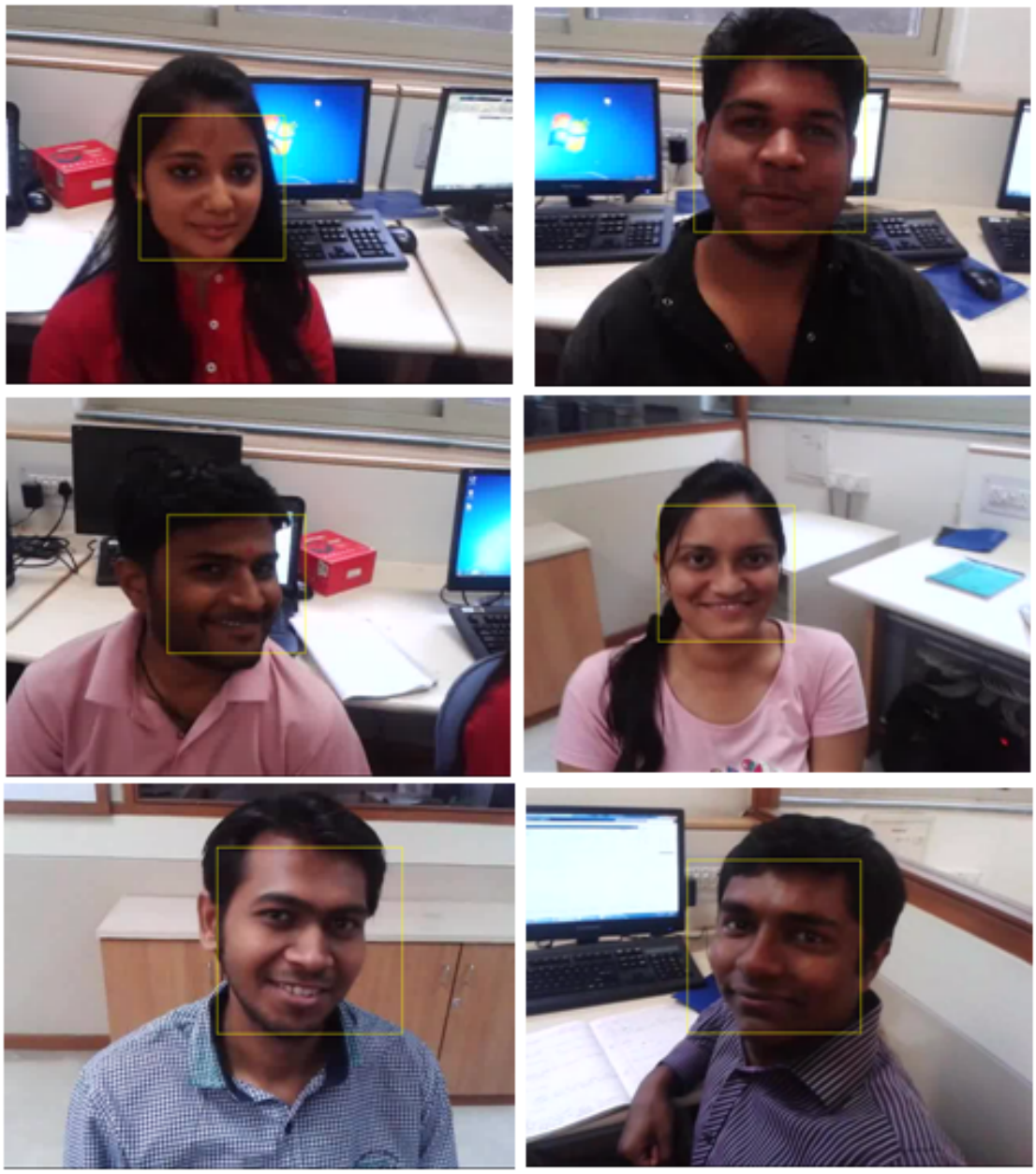

Figure 5.1: video frames with detected faces

Once the detection is done the faces are cropped from the video frames by their position and size, called subfaces, as shown in figure:

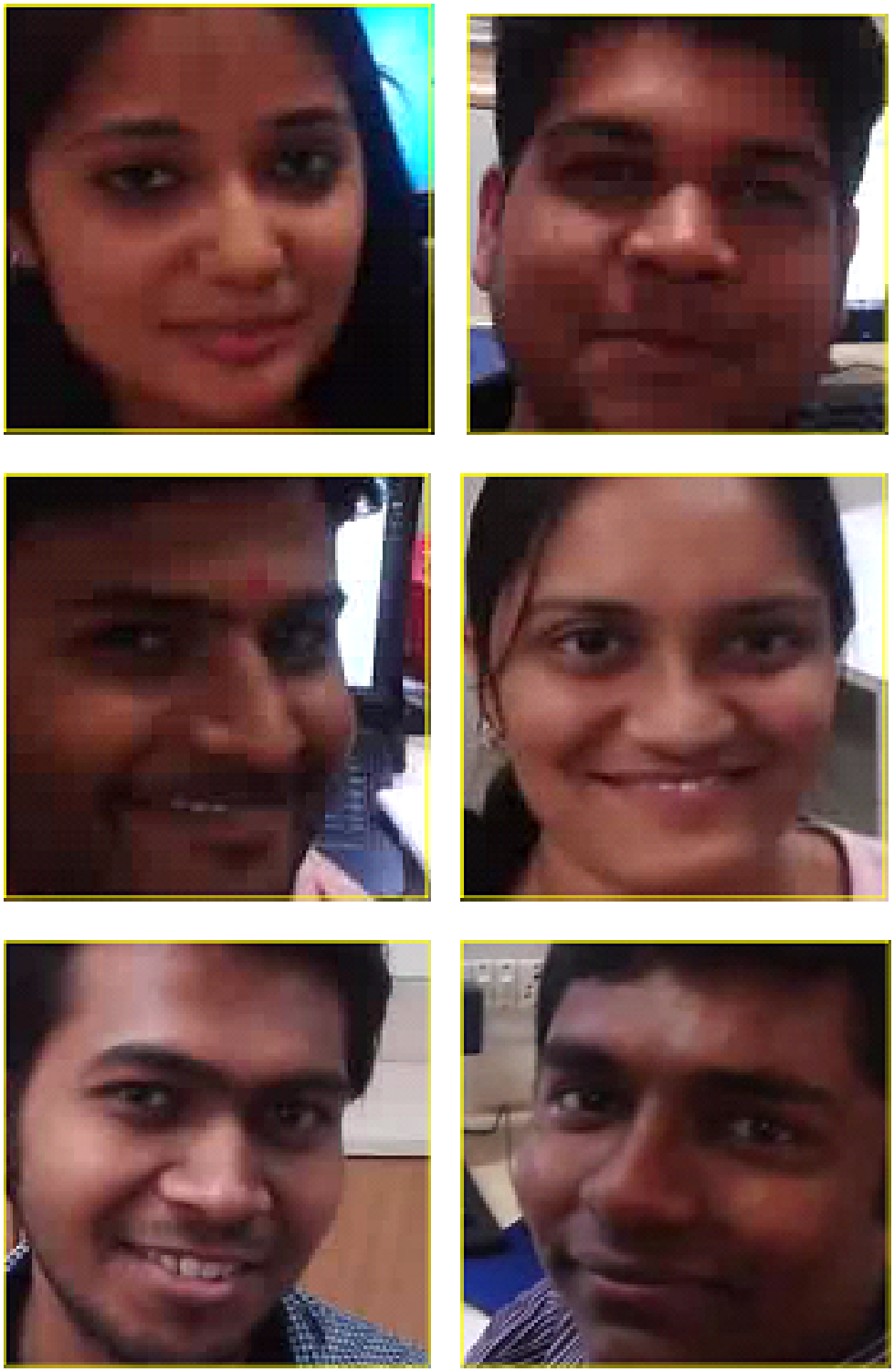

Figure 5.2: Subfaces

5.1.2 Dataset of unwanted users

The dataset contains the image of the unwanted user or terrorist, to whom we don't want to allow to enter in our sensitive areas. We consider that any new entry to the dataset of any unwanted person is made separately. Since the feature extraction is done with grayscale images. The coloured image frame face is first converted to greyscale and them matched to dataset, therefore the images in dataset should be in greyscale.

Figure 5.3: set of unwanted faces

5.1.3 Face matching: Orignal video vs set of unwanted faces

The faces, detected and generated from the original input video, are fed to feature extraction tecnique, which matches the feature in face detected with the feature of face in dataset of unwanted user. If the features are more than a certain threshold level then it is considered as face matched and the image which is matched from dataset is shown along with its name. The alarm is raised in this situation which alerts all the security guard about that person so that it is easy to catch the person. The matched faces are shown below:

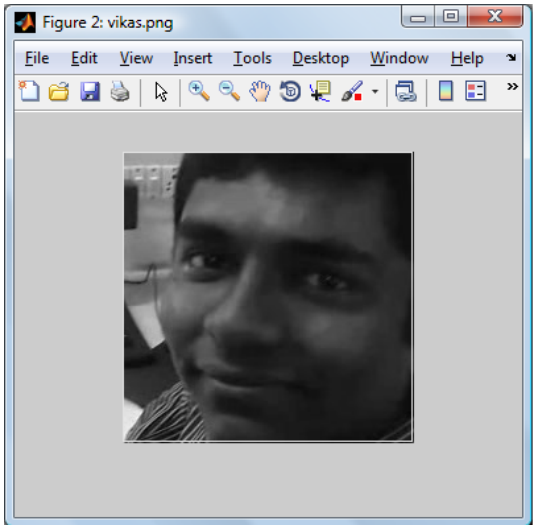

Figure 5.4: Unwanted Face matched with terrorist1

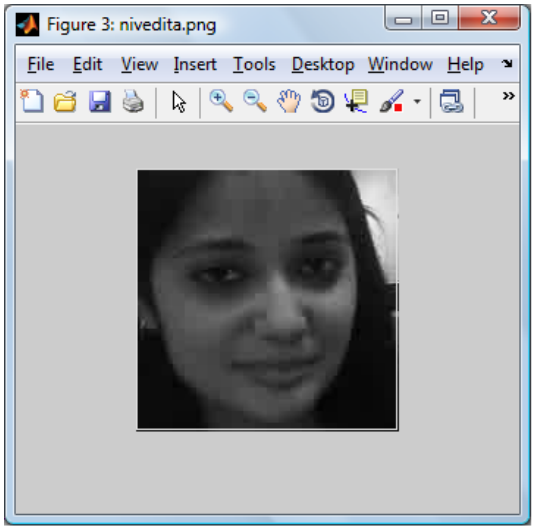

Figure 5.5: Unwanted Face matched with terrorist2

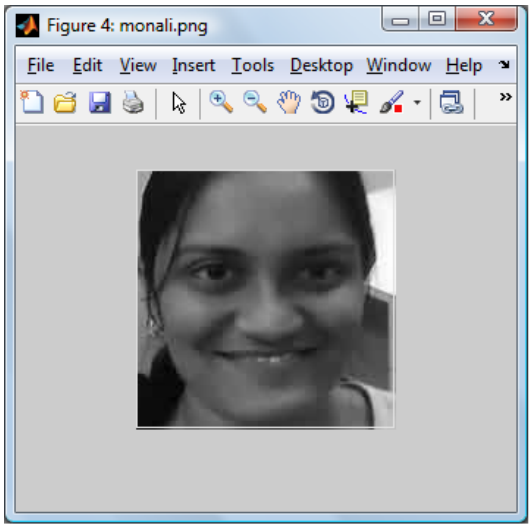

Figure 5.6: Unwanted Face matched with terrorist3

Chapter 6

Conclusions and Future Work

6.1 Conclusion

We have developed a Smart Surveillence System, where a camera is stationary and continuousl recording a video, this video processd further for frame generation, face localization, positioning, aligning, generating subfaces and then extracting features to match with the unwanted faces using any feature matching technique. If any faces from the database of unwanted person, matches then the system raise an alarm and for non-matching faces there is no action. This raised alarm alerts to the security person that an unwanted parson is in range or entereing the system. We can use these system at any sensitive places like Airport, shopping mall, railway staion, any tourist place etc.

6.2 Future Work

Currently our system takes too much time to match the faces of unwanted users. We are taking only some frames from the video and processing them for face detection and recognition. Since we are not taking all the frames so any unwanted face in skipped frame appears, will not be detected and our system will not be safe. Therefore our future purspective we will enhance the algorithm to detect and match faces faster than the original. Also we will try to take all the frames i.e. each frame of the video. So that we will detect unwanted face in all the frames, so that none of the face skips.

Bibliography

[1] L.S. Oliveira, D.L. Borges, F.B. Vidal, L. Chang, " A fast eye localization and verification method to improve face matching in surveillance videos,"IEEE International Conference on Systems, Man, and Cybernetics (SMC), pp.840-845, October 2012.

[2] Technological and Commercial Intelligence Report,Aude-Emmanuelle Fleurant,CRIM, Technople Defence and Security,April 8, 2009,"Intelligent Video Surveillance: Promises and Challenges"

[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] 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

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

[6] 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

[7] 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

[8] 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.

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

[10] W. Zhao, R. Chellappa, A. Rosenfeld, and P. Phillips. Face recognition: A literature survey. ACM Computing Surveys, pages 399-458, 2003.

[11] M.-H. Yang, D. Kriegman, and N. Ahuja. Detecting faces in images: A survey. IEEE Transactions on Pattern Analysis and Machine Intel- ligence, 24(1):34-58, January 2002.

[12] R. Brunelli and T. Poggio. Face recognition: Features versus templates. IEEE Transactions on Pattern Analysis and Machine Intelli- gence, 15(10):1042-1052, October 1993.

[13] S. K. Singh, D. S. Chauhan, M. Vatsa, and R. Singh. A robust skin color based face detection algorithm. Tamkang Journal of Science and Engineering, 6(4):227-234, 2003.

[14] CHETAN BALLUR,SHYLAJA S S, "APPLICATION OF LOCAL BINARY PATTERN AND PRINCIPAL COMPONENT ANALYSIS FOR FACE RECOGNITION",International Journal of Electrical, Electronics and Data Communication,P.E.S.I.T, Bangalore.

[15] Mallikarjuna Rao G,Vijaya Kumari G, Babu G R, Rajesh V, "Fast Local Binary Patterns for Efficient Face Recognition", Internat. J. of Sci. and Eng. Vol. 2(2):22-26, Dec. 2011.

[16] T.Ahonen, A.Hadid and M. Pietikainen. Face recognition with Local Binary Patterns. Machine Vision Group, University of Oulu, Finland,2004.

[17] Etemad, K., Chellappa, R.: Discriminant analysis for recognition of human face images. Journal of the Optical Society of America 14 1997.