HUMAN DETECTION IN SURVEILLANCE VIDEO

The Project submitted to

Sant Gadgebaba Amravati University, Amravati

Towards partial fulfilment of the Degree of

Bachelor of Engineering

In

Information Technology

Guided by

Prof. Ms. P. P. Bute

Submitted by

Mr. Ajay Dandge
Mr. Ajit Patil
Mr. Hrushikesh Shukla
Mr. Mayur Nehare

DEPARTMENT OF INFORMATION TECHNOLOGY
SHRI SANT GAJANAN MAHARAJ COLLEGE OF ENGINEERING, SHEGAON (M.S.)

2022- 2023

SHRI SANT GAJANAN MAHARAJ COLLEGE OF ENGINEERING, SHEGAON

2022-2023

CERTIFICATE

This is to certify that **Mr Ajay Dandge, Mr. Ajit Patil, Mr. Hrushikesh Shukla, Mr. Mayur Nehare,** students of final year B.E. (Information Technology) in the year 2022-2023 of the Information Technology Department of this institute have completed the project work entitled "**Human Detection in Surveillance Video**" based on syllabus and has submitted a satisfactory account of his/her work in this report which is recommended for the partial fulfilment of the degree of Bachelor of Engineering in Information Technology.

Name of Guide
Prof. Ms. P. P. Bute

Dr. A. S. Manekar
Head of the Department
SSGMCE, Shegaon

Dr. S. B. Somani
Principal
SSGMCE, Shegaon

SHRI SANT GAJANAN MAHARAJ COLLEGE OF ENGINEERING, SHEGAON

2022-2023

CERTIFICATE

This is to certify that the project work entitled "**Human Detection in Surveillance Video**" submitted by **Mr Ajay Dandge, Mr. Ajit Patil, Mr. Hrushikesh Shukla, Mr. Mayur Nehare**, students of final year B.E. (Information Technology) in the year 2022-2023 of the Information Technology Department of this institute, is a satisfactory account of his work based on the syllabus which is approved for the award of the degree of Bachelor of Engineering in Information Technology.

Internal Examiner

Date:

External Examiner

Date:

ACKNOWLEDGEMENT

The real spirit of achieving goals through the way of excellence and lustrous discipline. We would have never succeeded in completing our task without the cooperation, encouragement and help provided to us by various personalities.

We would like to take this opportunity to express our heartfelt thanks to our guide ***Prof. Ms. P. P. Bute,*** *for her esteemed guidance and encouragement, especially through difficult times. Her suggestions broaden our vision and guide us to succeed in this work. We are also very grateful for her guidance and comments while studying part of our project and learning many things under her leadership.*

We would also like to extend our sincere thanks to ***Prof. F. I. Khandwani****, Project-In-Charge for his valuable support and feedback during the entire course of the project.*

We also extend our thanks to ***Dr. A. S. Manekar****, Head of Information Technology Department, Shri Sant Gajanan Maharaj College of Engineering, Shegaon for providing us with a variety of opportunities and inspirations to gather professional knowledge and material that made us consistent performers.*

We also extend our thanks to ***Dr. S. B. Somani****, Principal, Shri Sant Gajanan Maharaj College of Engineering, Shegaon for providing us the infrastructure and facilities without which it was impossible to complete this work.*

Also, we would like to thank all teaching and non-teaching staff of the department for their encouragement, cooperation and help. Our greatest thanks to all those who wished us success, especially parents and friends.

Student Names

1. *Ajay Dandge*
2. *Ajit Patil*
3. *Hrushikesh Shukla*
4. *Mayur Nehare*

ABSTRACT

In recent years, there has been a lot of focus on the crucial task of human detection in surveillance footage. Security, public safety, traffic monitoring, and healthcare are just a few areas where it might be used. There are cameras installed in various public and private places to ensure safety and security. However, going through hours of surveillance footage can be a daunting and time-consuming task, and investigators may miss crucial information while analyzing the footage. Several processes are involved in the human detection process, including object detection, face detection, and tracking. This process can take hours, and investigators may miss important information while going through the footage. Another challenge is the lack of accuracy in manual analysis. Humans can get tired, lose focus, or miss crucial information while analyzing the footage. Moreover, humans can be biased and may not be able to identify specific faces or events accurately. Therefore, there is a need for a more accurate and efficient method of analyzing surveillance footage. Human Detection in Surveillance Video is a project that aims to provide a more efficient and accurate method of analyzing surveillance footage. The system can help reduce the time and effort required to go through hours of footage manually and can provide accurate and reliable results. The system can be used in various fields, including criminal investigations, identifying missing persons, and ensuring public safety.

Key Words: human detection, surveillance video, face detection

TABLE OF CONTENTS

List of Figures

1. INTRODUCTION

1.1 Preface

The use of surveillance cameras has increased significantly over the years. These cameras are installed in various public and private places to ensure safety and security. However, going through hours of surveillance footage can be a daunting and time-consuming task, and investigators may miss crucial information while analyzing the footage. Human Detection in Surveillance Video is a project that aims to reduce the human effort required during investigations and provide a more efficient and accurate method of analyzing surveillance footage.

One of the main challenges in analyzing surveillance footage is the large amount of data that needs to be processed. The footage from multiple cameras can generate terabytes of data that need to be analyzed manually. This process can take hours, and investigators may miss important information while going through the footage. Another challenge is the lack of accuracy in manual analysis. Humans can get tired, lose focus, or miss crucial information while analyzing the footage. Moreover, humans can be biased and may not be able to identify specific faces or events accurately. Therefore, there is a need for a more accurate and efficient method of analyzing surveillance footage.

The development of Human Detection in Surveillance Video is an important step in providing a more efficient and accurate method of analyzing surveillance footage. This system can help reduce the time and effort required to go through hours of footage manually and can provide accurate and reliable results. The system can be used in various fields, including criminal investigations, identifying suspects, monitoring employee behavior, and ensuring public safety. In criminal investigations, the system can help identify suspects, locate missing persons, and provide crucial evidence for solving crimes. In the workplace, the system can help monitor employee behavior, detect security breaches, and improve overall safety and security. In public places, the system can help detect potential threats and ensure public safety.

The implementation of Human Detection in Surveillance Video involves various processes. The first step is to gather the necessary hardware and software for the system. The system requires a computer with a high-performance processor, RAM, and a graphics card. It also requires a Face-Recognition Library, which can detect and recognize faces from surveillance footage. The next step is to develop the software for the system. The software should be able to analyze the surveillance footage and detect and recognize faces. The software should also be able to match the faces with a database of known faces and provide the timestamp and details of the wanted person.

The system should also have a user-friendly interface that allows users to upload pictures or videos, use a webcam to model CCTV footage, and add details of persons to be detected. The admin should be able to view the details of persons and mark them as wanted whom they want to detect.

Machine learning is a subset of artificial intelligence that involves the development of algorithms that can learn from data. Machine learning can be used in Human Detection in Surveillance Video to improve the accuracy and efficiency of the system. Machine learning algorithms can be trained on a dataset of known faces to improve the accuracy of face recognition. The algorithms can also be used to classify datasets and use supervised machine learning algorithms to improve crop productivity.

Human Detection in Surveillance Video is a project that aims to provide a more efficient and accurate method of analyzing surveillance footage. The system can help reduce the time and effort required to go through hours of footage manually and can provide accurate and reliable results. The system can be used in various fields, including criminal investigations, identifying suspects, monitoring employee behavior, and ensuring public safety.

The development of this project involves various processes, including gathering the necessary hardware and software, developing the software, and implementing machine learning algorithms. The system also provides user-friendly interface that allows users to perform functionalities provided by the system.

1.2 Statement of Problem

The increasing need for safety and security in public spaces has led to the implementation of surveillance systems. These systems are designed to monitor activities and detect potential threats to public safety. However, with the vast amount of data being collected by surveillance cameras, it becomes difficult for human operators to effectively monitor every feed. This is where the need for a human detection system arises. The accurate detection of humans in surveillance videos can help reduce the workload of human operators and provide early detection of potential security threats. The challenge lies in developing the system that can accurately detect humans in varying lighting and environmental conditions while minimizing false positives.

- **Solution Requirement**

We analyzed the problem statement and identified the need for a robust human detection model that can accurately detect humans in surveillance videos to ensure security and safety. We conducted a thorough literature review to identify the most effective algorithms and techniques for human detection in video data. After training and testing the model, we evaluated its performance to ensure its effectiveness in detecting humans in surveillance videos.

- **Solution Constraints**

We analyzed the solution in terms of cost, speed of processing, requirements, level of expertise, and availability of equipment's.

1.3 Objectives of the project

The objectives of this project are:

- To develop a system that can detect and recognize faces in surveillance footage.
- To reduce the time and effort required to go through hours of footage manually.
- To provide investigators with a tool that can automate the process of identifying a specific person or event in surveillance footage.

- To develop a web-based interface that is user-friendly and can be used by investigators with varying levels of technical expertise.
- To use open-source technologies to develop the system, making it scalable and cost-effective.
- To improve the efficiency of investigations by automating the process of identifying a person in surveillance footage.
- To help investigators focus on other aspects of their investigation, leading to quicker and more efficient investigations.

1.4 Scope and limitation of the Project

Human Detection in Surveillance Video will help in reducing human efforts required during investigations by providing the timestamp and details of the wanted person by showing the exact frame when he/she is first time detected. Our solution will be a web-based platform for the user to upload the surveillance video. The platform can be used by security personnel and law enforcement agencies to enhance public safety and prevent criminal activities. We can also consider scaling up the project to develop a browser plugin for automatic human detection in surveillance videos. Our project aims to develop an accurate human detection model for surveillance videos, which can help in ensuring public safety and preventing criminal activities. We have analyzed the problem statement and found the feasibility of the solution.

Limitations:

While the project has a wide scope and potential impact, there are some limitations that should be considered, including:

- The accuracy of the system will depend on the quality of the images or videos captured by the CCTV cameras.
- The system may not work properly in low light or poor visibility conditions.
- The system may not be able to detect faces accurately if the person is wearing a mask or their face is covered.
- The system may not be able to detect people who are not in its database or have not been marked as wanted.

1.5 Organization of the Project

The project is organized as follows:

1. Chapter 1 gives Introduction about the project.
2. Chapter 2 gives Literature survey of the project.
3. Chapter 3 provides analysis of project.
4. Chapter 4 provides design phase of project.
5. Chapter 5 provides how project is implemented.
6. Chapter 6 gives conclusion with future scope of the project

2. LITERATURE SURVEY

Paper 1 : "Human Identification Recognition In Surveillance Videos" by Kai Jin, Xuemei Xie , Fangyu Wang, Xiao Han, Guangming Shi [1]

Despite having superior performance, users still requires to cooperate with the camera while using facial-recognition. Due to the blurred and warped settings in surveillance footage, human identify recognition works poorly. In this study, authors provide a fresh approach to identify human identity by combining global and local structural data. First, in order to enhance the recognition of faces with occlusions and facial deformities, researchers integrate pedestrian-detection and tracking with face recognition. Additionally, selective-recognition method based on person's trajectory that may be used to recognize a person approaching a webcam are offered. Compared to techniques that just use facial information, our strategy performs better. Human identification recognition performs poorly in surveillance videos because of the deformed and occluded conditions. In this paper, we propose a new method to recognize human identification using global and local structural information. Firstly, author combine pedestrian detection and tracking with face recognition, in order to improve the recognition performance of occlusion and deformed faces. Secondly, author propose a selective recognition algorithm based on pedestrian trajectory, which is used to identify the pedestrian walking towards the webcam. Finally, author conduct experiments on our own human identification dataset, which contains 93 challenging video sequences captured in the corridor. This method achieves better performance than methods only using face information.

Paper 2 : "Human Detection and Tracking for Video Surveillance: A Cognitive Science Approach" by Vandit Gajjar, Ayesha Gurnani, Yash Khandhediya [2]

Video surveillance is an important aspect of security in many domains, such as public places, offices, and homes. The ability to detect and track human subjects in real-time is essential for ensuring effective surveillance and security. In this paper, the authors propose a cognitive science-based approach to human detection and tracking for video surveillance. The proposed approach is based on the concept of attentional control,

which is a fundamental cognitive process involved in visual perception and attention. The authors use a combination of computer vision techniques and cognitive science principles to design an attentional control model that can detect and track human subjects in a video stream. The model incorporates three main components: feature extraction, attentional control, and tracking. The feature extraction component uses convolutional neural networks (CNNs) to extract relevant features from the video frames. The attentional control component employs a saliency map to guide the model's attention to the most relevant regions of the video stream. The tracking component uses a Kalman filter to estimate the position and velocity of the detected human subjects.
The authors evaluate the proposed approach on a dataset of video surveillance footage and compare it with state-of-the-art human detection and tracking techniques. The results show that their approach outperforms existing methods in terms of accuracy, speed, and robustness. Overall, the authors' cognitive science-based approach provides a promising solution for human detection and tracking in video surveillance systems. The attentional control model can adapt to changes in the environment and handle complex scenarios, making it suitable for real-world applications.

Paper 3 : "Real-Time Criminal Identification System Based On Face Recognition" by Mr.R.Prashanth Kumar, Abdul Majeed, Farhan Pasha, A Sujith [3]

This facial recognition-based real-time criminal detection system uses a facial recognition technology that is completely automated. For face identification and recognition, OpenCV LBPH (Local Binary Pattern Histograms) Algorithms and the Haar feature-based cascade classifier are both employed. This technology will be able to recognise faces in real time and automatically detect faces. It's still difficult to pinpoint the face with accuracy. The Viola-Jones framework has been widely utilised by academics to identify faces and other objects in a picture. Public communities like OpenCV share face detection classifiers. The proposed system is designed to work in real-time and can identify criminals in a matter of seconds. The system is also scalable and can be easily integrated with existing surveillance infrastructure. The authors evaluate the proposed system using a dataset of surveillance footage and compare it with existing face recognition systems. The results show that their system outperforms existing methods in terms of accuracy and speed. Overall, the authors' real-time

criminal identification system provides a powerful tool for law enforcement agencies to identify and apprehend criminals quickly and efficiently. The system has the potential to significantly improve public safety and help prevent crime.

Paper 4 : "Human Identification Using Human Body Features Extraction" by Martino C. Khuangga & Dwi H. Widyantoro [4]

Author proposes a human-identification-system that uses human body feature extraction and can monitor a room's occupants' presence. The main goal of this system is to identify people entering a specific area or room without requiring them to take any specific action, such as showing their identification card or scanning their fingerprints. The system uses a camera as an input device and chooses features on the human body because they are often simpler to identify and serve as a strong identification to label people who enter a place. The proposed system uses two primary processes. The first process is the detection of when someone enters a room. The system does this by analyzing the video feed from the camera and looking for the presence of a person in the frame. Once a person is detected, the system extracts the features from their body and stores them in a database. The second process is the detection of when the same individual leaves the room. The system achieves this by constantly monitoring the video feed from the camera and looking for changes in the number of people in the room. When a person is no longer detected in the room, the system compares their features with those stored in the database. If a match is found, the system logs the time and date when the person left the room. To achieve accurate and reliable human body feature extraction, the system utilizes various image processing methods, including HOG descriptor-based person recognition, HSV colour conversion, and template matching. The HOG descriptor-based person recognition method is used to identify people based on their body shape and texture, while the HSV colour conversion is used to extract colour information from the video frames. The template matching method is used to locate specific body parts, such as the head or the torso. The authors evaluate the proposed system's performance using a dataset of video footage and compare it with existing human identification systems. The results show that their system can accurately identify people in real-time and achieve a high recognition rate. In summary, the proposed human identification system using human body feature extraction offers a

non-intrusive and efficient solution for identifying people entering and leaving a specific area or room. The system's ability to operate without requiring any specific action from the occupants makes it suitable for various applications, including surveillance, security, and access control.

Paper 5 : R. Rathi, M. Choudhary & B. Chandra, "An Application of Face Recognition System using Image Processing and Neural Networks", International Journal Computer Technology Application, 3:1, (2012), pp. 45-49. [5]

In recent years face recognition has received substantial attention from both research communities and the market, but still remained very challenging in real applications. A lot of face recognition algorithms, along with their medications, have been developed during the past decades. A number of typical algorithms are presented. In this paper, we propose to label a Self-Organizing Map (SOM) to measure image similarity. To manage this goal, we feed Facial images associated to the regions of interest into the neural network. At the end of the learning step, each neural unit is tuned to a particular Facial image prototype. Facial recognition is then performed by a probabilistic decision rule. This scheme offers very promising results for face identification dealing with illumination variation and facial poses and expressions. This paper presents a novel Self-Organizing Map (SOM) for face recognition. The SOM method is trained on images from one database. The novelty of this work comes from the integration of A facial recognition system is a computer application for automatically identifying or verifying a person from a digital image or a video frame from a video source. One of the way is to do this is by comparing selected facial features from the image and a facial database.

Paper 6 : "PCA in the context of Face Recognition with the Image Enlargement Techniques" by Mohammed Kabiru Halidu, Pooneh Bagheri-Zadeh, Akbar Sheikh-Akbari, Reinhold Behringer [6]

Description :

Face recognition has become a field of interest in many applications such as security and entertainments. In surveillance system, the quality of the recoded footage is

sometimes insufficient due to the distance and angle of the camera from the scene. This causes the object of interest, e.g. the face of a person in the scene to be of low resolution, which increases the difficulty in recognition process. Image resolution enhancement is a potential solution for enlarging low-resolution images for real time face recognition. However, the optimal performance of face recognition techniques when various image enlargement methods have been applied to them has not been investigated. In this research, the performance of PCA based face recognition method, with the three most well-known image enlargement techniques (Nearest Neighbour, Bilinear, Bicubic) is investigated. First, an input image is down sampled to six different resolutions. The down-sampled image is then enlarged to its original size using the three named image enlargement techniques. The enlarged image is then input to a PCA face recognition system for the recognition process. The simulation results using images from the SCFace database show that PCA based face recognition illustrates superior results when input images enlarged using Nearest Neighbour technique, while the performance of Bicubic and Bilinear techniques is slightly lower than Nearest Neighbour method.

Paper 7 : "An introduction to biometric recognition" by A. K. Jain, Arun A. Ross, Nandakumar, K. [7]

Description :

A wide variety of systems requires reliable personal recognition schemes to either confirm or determine the identity of an individual requesting their services. The purpose of such schemes is to ensure that the rendered services are accessed only by a legitimate user and no one else. Examples of such applications include secure access to buildings, computer systems, laptops, cellular phones, and ATMs. In the absence of robust personal recognition schemes, these systems are vulnerable to the wiles of an impostor. Biometric recognition, or, simply, biometrics, refers to the automatic recognition of individuals based on their physiological and/or behavioral characteristics. By using biometrics, it is possible to confirm or establish an individual's identity based on "who she is", rather than by "what she possesses" (e.g., an ID card) or "what she remembers" (e.g., a password). We give a brief overview of the field of biometrics and summarize some of its advantages, disadvantages, strengths, limitations, and related privacy concerns.

Paper 8 : "Face Recognition Based on Symmetrical Half-Join Method using Stereo Vision Camera" by Edy Winarno, Agus Harjoko, Aniati Murni Arymurthy, Edi Winarko [8]

Description :

The main problem in face recognition system based on half-face pattern is how to anticipate poses and illuminance variations to improve recognition rate. To solve this problem, we can use two lenses on stereo vision camera in face recognition system. Stereo vision camera has left and right lenses that can be used to produce a 2D image of each lens. Stereo vision camera in face recognition has capability to produce two of 2D face images with a different angle. Both angle of the face image will produce a detailed image of the face and better lighting levels on each of the left and right lenses. In this study, we proposed a face recognition technique, using 2 lens on a stereo vision camera namely symmetrical half-join. Symmetrical half-join is a method of normalizing the image of the face detection on each of the left and right lenses in stereo vision camera, then cropping and merging at each image. Tests on face recognition rate based on the variety of poses and variations in illumination shows that the symmetrical half-join method is able to provide a high accuracy of face recognition and can anticipate variations in given pose and illumination variations. The proposed model is able to produce 86%-97% recognition rate on a variety of poses and variations in angles between 0°-22.5°. The variation of illuminance measured using a lux meter can result in 90%-100% recognition rate for the category of at least dim lighting levels (above 10 lux)

Paper 9: "Rethinking the Inception Architecture for Computer Vision", by Christian Szegedy, Vincent Vanhoucke, Sergey Ioffe, Jon Shlens [9]

Description: Convolutional networks are at the core of most state of-the-art computer vision solutions for a wide variety of tasks. Since 2014 very deep convolutional networks started to become mainstream, yielding substantial gains in various benchmarks. Although increased model size and computational cost tend to translate to immediate quality gains for most tasks (as long as enough labeled data is provided for training), computational efficiency and low parameter count are still enabling factors

for various use cases such as mobile vision and big-data scenarios. Here ways to scale up networks in ways that aim at utilizing the added computation as efficiently as possible by suitably factorized convolutions and aggressive regularization are explored. This methods were benchmarked on the ILSVRC 2012 classification challenge validation set demonstrate substantial gains over the state of the art: 21.2% top-1 and 5.6% top-5 error for single frame evaluation using a network with a computational cost of 5 billion multiply-adds per inference and with using less than 25 million parameters. With an ensemble of 4 models and multi-crop evaluation, report 3.5% top-5 error and 17.3% top-1 error on the validation set and 3.6% top-5 error on the official test set. Several design principles to scale up convolutional networks and studied them in the context of the Inception architecture are provided. This guidance can lead to high performance vision networks that have a relatively modest computation cost compared to simpler, more monolithic architectures. This highest quality version of Inception-v2 reaches 21.2%, top-1 and 5.6% top-5 error for single crop evaluation on the ILSVR 2012 classification, setting a new state of the art. The combination of lower parameter count and additional regularization with batch-normalized auxiliary classifiers and label-smoothing allows for training high quality networks on relatively modest sized training sets.

Paper 10 : "Introduction to face recognition and evaluation of algorithm performance" by G.H. Givens, J.R. Beveridge, P.J. Phillips, B. Draper, Y.M. Lui, D. Bolme [10]

Description :

The field of biometric face recognition blends methods from computer science, engineering and statistics, however statistical reasoning has been applied predominantly in the design of recognition algorithms. A new opportunity for the application of statistical methods is driven by growing interest in biometric performance evaluation. Methods for performance evaluation seek to identify, compare and interpret how characteristics of subjects, the environment and images are associated with the performance of recognition algorithms. Some central topics in face recognition are reviewed for background and several examples of recognition algorithms are given. One approach to the evaluation problem is then illustrated with a generalized linear mixed model analysis of the Good, Bad, and Ugly Face Challenge, a pre-eminent face

recognition dataset used to test state-of-the-art stillimage face recognition algorithms. Findings include that (i) between-subject variation is the dominant source of verification heterogeneity when algorithm performance is good, and (ii) many covariate effects on verification performance are 'universal' across easy, medium and hard verification tasks. Although the design and evaluation of face recognition algorithms draw upon some familiar statistical ideas in multivariate statistics, dimension reduction, classification, clustering, binary response data, generalized linear models and random effects, the field also presents some unique features and challenges. Opportunities abound for innovative statistical work in this new field. The results presented here illustrate how statistical models can be applied in a manner to help algorithm developers and users understand – and hopefully improve – recognition performance. It is clear that the statistical toolbox contains many more applicable methods for development and evaluation. Performance evaluation is a particularly fertile topic for statistical attention due to burgeoning interest from the biometric community and the fact that a predictor–response formulation is one natural way to pose key questions using available datasets. More broadly, opportunities for statistical research in face recognition are abundant because many aspects of recognition algorithms, the structure of the identification and verification tasks they face, and their performance have only recently begun to be posed explicitly in a statistical context.

Paper 11 : "A Review Of Face Recognition Methods" By M. Parisa Beham, S. Mohamed Mansoor Roomi [11]

Description :

Face recognition has become more signifcant and relevant in recent years owing to it potential applications. Since the faces are highly dynamic and pose more issues and challenges to solve, researchers in the domain of pattern recognition, computer vision and arti¯cial intelligence have proposed many solutions to reduce such di±culties so as to improve the robustness and recognition accuracy. As many approaches have been proposed, e®orts are also put in to provide an extensive survey of the methods developed over the years. The objective of this paper is to provide a survey of face recognition papers that appeared in the literature over the past decade under all severe conditions that were not discussed in the previous survey and to categorize them into

meaningful approaches, viz. appearance based, feature based and soft computing based. A comparative study of merits and demerits of these approaches have been presented. Appearance-based approaches attempt to recognize faces using global representations, i.e. descriptions based on the entire image rather than on local features of the face. The main advantage of the holistic approaches is that they do not obliterate any of the information in the images by focussing on only limited regions. However, as mentioned above, this same property is their utmost drawback, too, since most of these approaches start out with the basic assumption that all the pixels in the image are equally important.

Paper 12 : "Real-time surveillance of people and their activities", I. Haritaoglu, D. Harwood and L. S. Davis [12]

Description :

Paper describes the real time visual surveillance system for detecting and tracking multiple people and monitoring their activities in an outdoor environment. It operates on monocular gray-scale video imagery, or on video imagery from an infrared camera. W/sup 4/ employs a combination of shape analysis and tracking to locate people and their parts (head, hands, feet, torso) and to create models of people's appearance so that they can be tracked through interactions such as occlusions. It can determine whether a foreground region contains multiple people and can segment the region into its constituent people and track them. W/sup 4/ can also determine whether people are carrying objects, and can segment objects from their silhouettes, and construct appearance models for them so they can be identified in subsequent frames. W/sup 4/ can recognize events between people and objects, such as depositing an object, exchanging bags, or removing an object. It runs at 25 Hz for 320/spl times/240 resolution images on a 400 MHz dual-Pentium II PC.

Paper 13 : "A Comparative Analysis of Image Interpolation Algorithms" by Mr. Pankaj S. Parsania1 , Dr. Paresh V. Virparia [13]

Description :

Image interpolation is a term used for image processing, but is often used with different terminologies in literature like image scaling, image resampling and image resize.

There are many algorithms currently in use for the resizing of digital images. Most of them attempt to reproduce a visually attractive replica of the original. Now a days the technology for smaller display area for viewing on varieties of device, image size is generally down sampled (or sub sampled or decreased) in order to produce thumbnails. Up sampling an image (or enlarging or interpolating) is most common for viewing on large display size monitor or television. While enlarging an image it is not possible to discover any more information in the image than already exists, and image quality inevitably suffers. However, there are several methods of increasing the number of pixels that an image contains, which can be created out of original pixels. These methods are often termed as image interpolation algorithms. Image interpolation is the most basic requirement for many image processing task such as computer graphics, gaming, medical image processing, virtualization, camera surveillance and quality control. Image interpolation is a technique used in resizing images. To resize an image, every pixel in the new image must be mapped back to a location in the old image in order to calculate a value of new pixel. There are many algorithms available for determining new value of the pixel, most of which involve some form of interpolation among the nearest pixels in the old image. In this paper, we used Nearest-neighbor, Bilinear, Bicubic, Bicubic B-spline, Catmull-Rom, Mitchell Netravali and Lanzcos of order three algorithms for image interpolation. Each algorithms generates varies artifact such as aliasing, blurring and moiré.

Paper 14 : "An improved ORB feature extraction and matching algorithm based on affine transformation" By Zhifei Wang, Zhipeng Li, Lan Cheng, Gaowei Yan [14]

Description :

As the key technology of image processing, image feature extraction and matching are widely used in face recognition, image stitching, and visual SLAM. Among them, ORB algorithm is widely adopted because of its advantage in real-time processing. However, the matching accuracy of the feature points extracted by ORB algorithm is still a concern for further applications. proposed an improved ORB algorithm, which reduced the Hamming distance of the descriptor from the same feature points by extracting the descriptor stable bits under different affine transformations. Furthermore, after feature matching, we employed an improved F-SORT algorithm to purify the matching

features. The experimental results show that the improved ORB algorithm can effectively improve the matching accuracy and increase the number of correct matches. The matching accuracy was further improved by applying an improved F-SORT algorithm. The proposed algorithm is suitable for real-time occasions such as visual SLAM. However, the performance of the F-SORT algorithm is closely related to the accuracy of feature matching, and performs poorly when the matching accuracy is low. Further study will be further done to address this issue.To address the problem, an improved ORB algorithm based on affine transformation was proposed. After FAST feature points are detected, the descriptors under different affine transformations are extracted from the feature points, and then the stable bits of the descriptor are extracted for feature matching. To further improve the accuracy of feature matching, an improved F-SORT algorithm is employed after feature matching. The improved F-SORT algorithm first groups the matched features in sequence order, and then further refines the matched feature pairs based on the angle, scale, and distance of the feature points. Experimental results verify the effectiveness of the algorithm regarding time efficiency and matching accuracy. proposed an improved ORB algorithm, which reduced the Hamming distance of the descriptor from the same feature points by extracting the descriptor stable bits under different affine transformations. Furthermore, after feature matching, we employed an improved F-SORT algorithm to purify the matching features. The experimental results show that the improved ORB algorithm can effectively improve the matching accuracy and increase the number of correct matches. The matching accuracy was further improved by applying an improved F-SORT algorithm. The proposed algorithm is suitable for real-time occasions such as visual SLAM. However, the performance of the F-SORT algorithm is closely related to the accuracy of feature matching, and performs poorly when the matching accuracy is low. Further study will be further done to address this issue.

Paper 15 : "Evaluation of Local Detectors and Descriptors for Fast Feature Matching" By Ondrej Miksik CMP and Krystian Mikolajczyk [15]

Description :

Local feature detectors and descriptors are widely used in many computer vision applications and various methods have been proposed during the past decade. There

have been a number of evaluations focused on various aspects of local features, matching accuracy in particular, however there has been no comparisons considering the accuracy and speed trade-offs of recent extractors such as BRIEF, BRISK, ORB, MRRID, MROGH and LIOP. This paper provides a performance evaluation of recent feature detectors and compares their matching precision and speed in randomized kd trees setup as well as an evaluation of binary descriptors with efficient computation of Hamming distance. This paper reports the matching precision and speed of recent feature detectors and descriptors in multiple randomized kd-tree setup and Hamming distance bruteforce search. The recently proposed real valued descriptors such as LIOP, MRRID, MROGH outperform state of-the-art descriptors SURF and SIFT in both, precision and recall, although their efficiency is very low. Similarly, binary descriptors provide a very efficient technique for time-constrained applications with good matching accuracy. Another advantages of binary descriptors are very fast extraction times and very low memory requirements, since each descriptor has only 32 (BRIEF, ORB) or 64 (BRISK) bytes, respectively. These descriptors provide comparable precision/recall results with SURF and SIFT.

Paper 16 : "ORB: an efficient alternative to SIFT or SURF" by Ethan Rublee, Vincent Rabaud , Kurt Konolige and Gary Bradski [16]

Description :

Feature matching is at the base of many computer vision problems, such as object recognition or structure from motion. Current methods rely on costly descriptors for detection and matching. In this paper, we propose a very fast binary descriptor based on BRIEF, called ORB, which is rotation invariant and resistant to noise. We demonstrate through experiments how ORB is at two orders of magnitude faster than SIFT, while performing as well in many situations. The efficiency is tested on several real-world applications, including object detection and patch-tracking on a smart phone. In this paper, we propose a computationally-efficient replacement to SIFT that has similar matching performance, is less affected by image noise, and is capable of being used for real-time performance. Our main motivation is to enhance many common image-matching applications, e.g., to enable low-power devices without GPU acceleration to perform panorama stitching and patch tracking, and to reduce the time

for feature-based object detection on standard PCs. Our descriptor performs as well as SIFT on these tasks (and better than SURF), while being almost two orders of magnitude faster. In this paper, we have defined a new oriented descriptor, ORB, and demonstrated its performance and efficiency relative to other popular features. The investigation of variance under orientation was critical in constructing ORB and de-correlating its components, in order to get good performance in nearest-neighbor applications. We have also contributed a BSD licensed implementation of ORB to the community, via OpenCV 2.3.

Paper 17 : "BRIEF: Binary Robust Independent Elementary Features" by Michael Calonder, Vincent Lepetit, Christoph Strecha, and Pascal Fua [17]

Description :

Feature point descriptors are now at the core of many Computer Vision technolo-gies, such as object recognition, 3D reconstruction, image retrieval, and cameralocalization. Since applications of these technologies have to handle ever moredata or to run on mobile devices with limited computational resources, there isa growing need for local descriptors that are fast to compute, fast to match, andmemory efficient.One way to speed up matching and reduce memory consumption is to workwith short descriptors. They can be obtained by applying dimensionality reduc-tion, such as PCA or LDA, to an original descriptor such as SIFT or SURF . For example, it was shown that floating point values of thedescriptor vector could be quantized using very few bits per value without loss ofrecognition performance. An even more drastic dimensionality reduction can beachieved by using hash functions that reduce SIFT descriptors to binary strings,as done. These strings represent binary descriptors whose similarity canbe measured by the Hamming distance. Authors show that it is highly discriminative even when using relatively few bits and can be computed using simple intensity difference tests. Furthermore, the descriptor similarity can be evaluated using the Hamming distance, which is very efficient to compute, instead of the L2 norm as is usually done. As a result, BRIEF is very fast both to build and to match. We compare it against SURF and U-SURF on standard benchmarks and show that it yields a similar or better recognition performance, while running in a fraction of the time required by either.

Paper 18 : "SURF: Speeded Up Robust Features" by H. Bay, A. Ess, T. Tuytelaars and L. V. Gool [18]

Description :

The task of finding point correspondences between two images of the same scene or object is part of many computer vision applications. Image registration, camera calibration, object recognition, and image retrieval are just a few. The search for discrete image point correspondences can be divided into three main steps. First, 'interest points' are selected at distinctive locations in the image, such as corners, blobs, and T-junctions. The most valuable property of an interest point detector is its repeatability. The repeatability expresses the reliability of a detector for finding the same physical interest points under different viewing conditions. Next, the neighbourhood of every interest point is represented by a feature vector. This descriptor has to be distinctive and at the same time robust to noise, detection displacements and geometric and photometric deformations. Finally, the descriptor vectors are matched between different images. The matching is based on a distance between the vectors, e.g. the Mahalanobis or Euclidean distance. The dimension of the descriptor has a direct impact on the time this takes, and less dimensions are desirable for fast interest point matching. However, lower dimensional feature vectors are in general less distinctive than their high-dimensional counterparts.much faster. This is achieved by relying on integral images for image convolutions; by building on the strengths of the leading existing detectors and descriptors (in casu, using a Hessian matrix-based measure for the detector, and a distribution-based descriptor); and by simplifying these methods to the essential. This leads to a combination of novel detection, description, and matching steps.

Paper 19 : "Machine Learning for High-Speed Corner Detection" by Edward Rosten, Thomas Drummond [19]

Description :

Where feature points are used in real-time frame-rate applications, a high-speed feature detector is necessary. Feature detectors such as SIFT (DoG), Harris and SUSAN are good methods which yield high quality features, however they are too computationally

intensive for use in real-time applications of any complexity. Here we show that machine learning can be used to derive a feature detector which can fully process live PAL video using less than 7% of the available processing time. By comparison neither the Harris detector (120%) nor the detection stage of SIFT (300%) can operate at full frame rate. Clearly a high-speed detector is of limited use if the features produced are unsuitable for downstream processing. In particular, the same scene viewed from two different positions should yield features which correspond to the same real-world 3D locations. Hence the second contribution of this paper is a comparison corner detectors based on this criterion applied to 3D scenes. This comparison supports a number of claims made elsewhere concerning existing corner detectors. Further, contrary to our initial expectations, we show that despite being principally constructed for speed, our detector significantly outperforms existing feature detectors according to this criterion.

Paper 20 : "You Only Look Once: Unified, Real-Time Object Detection" by Joseph Redmon, Santosh Divvala, Ross Girshick, Ali Farhadi [20]

Description: YOLO, a new approach to object detection. Prior work on object detection repurposes classifiers to perform detection. Instead, object detection is frame ed as a regression problem to spatially separated bounding boxes and associated class probabilities. A single neural network predicts bounding boxes and class probabilities directly from full images in one evaluation. Since the whole detection pipeline is a single network, it can be optimized end-to-end directly on detection performance. This unified architecture is extremely fast. This base YOLO model processes images in real-time at 45 frames per second. A smaller version of the network, Fast YOLO, processes an astounding 155 frames per second while still achieving double the mAP of other real-time detectors. Compared to state-of-the-art detection systems, YOLO makes more localization errors but is less likely to predict false positives on background. Finally, YOLO learns very general representations of objects. It outperforms other detection methods, including DPM and R-CNN, when generalizing from natural images to other domains like artwork. Current detection systems repurpose classifiers to perform detection. To detect an object, these systems take a classifier for that object and evaluate it at various locations and scales in a test image. Systems like deformable parts models (DPM) use a sliding window approach where the classifier is run at evenly spaced

locations over the entire image. Here object detection is reframed as a single regression problem, straight from image pixels to bounding box coordinates and class probabilities. Using this system, you only look once (YOLO) at an image to predict what objects are present and where they are

Paper 21 : "YOLO9000: Better, Faster, Stronger", by Joseph Redmon, Ali Farhadi [21]

Description: YOLOv2 and YOLO9000, real-time detection systems are introduced. YOLOv2 is state-of-the-art and faster than other detection systems across a variety of detection datasets. Furthermore, it can be run at a variety of image sizes to provide a smooth tradeoff between speed and accuracy. YOLO9000 is a real-time framework for detection more than 9000 object categories by jointly optimizing detection and classification. WordTree is used to combine data from various sources and this joint optimization technique to train simultaneously on ImageNet and COCO. YOLO9000 is a strong step towards closing the dataset size gap between detection and classification. Many of this techniques generalize outside of object detection. This WordTree representation of ImageNet offers a richer, more detailed output space for image classification. Dataset combination using hierarchical classification would be useful in the classification and segmentation domains. Training techniques like multi-scale training could provide benefit across a variety of visual tasks. For future work hope to use similar techniques for weakly supervised image segmentation. There is also plan to improve this detection results using more powerful matching strategies for assigning weak labels to classification data during training. Computer vision is blessed with an enormous amount of labelled data. Lets continue looking for ways to bring different sources and structures of data together to make stronger models of the visual world.

Paper 22 : "Real Time Object Detection and Tracking Using Deep Learning and OpenCV", by Chandan G, Ayush Jain, Harsh Jain, Mohana [22]

Description: Deep learning has gained a tremendous influence on how the world is adapting to Artificial Intelligence since past few years. Some of the popular object

detection algorithms are Region-based Convolutional Neural Networks (RCNN), FasterRCNN, Single Shot Detector (SSD) and You Only Look Once (YOLO). Amongst these, Faster-RCNN and SSD have better accuracy, while YOLO performs better when speed is given preference over accuracy. Deep learning combines SSD and Mobile Nets to perform efficient implementation of detection and tracking. This algorithm performs efficient object detection while not compromising on the performance. Objects are detected using SSD algorithm in real time scenarios. Additionally, SSD have shown results with considerable confidence level. Main Objective of SSD algorithm to detect various objects in real time video sequence and track them in real time. This model showed excellent detection and tracking results on the object trained and can further utilized in specific scenarios to detect, track and respond to the particular targeted objects in the video surveillance. Further extending the work to detect ammunition and guns in order to trigger alarm in case of terrorist attacks. The model can be deployed in CCTVs, drones and other surveillance devices to detect attacks on many places like schools, government offices and hospitals where arms are completely restricted.

Paper 23 : "Real Time Boundary Density based Moving Object Extraction", by Panca Mudjirahardjo and Hadi Suyono [23]

Description: The first task in motion detection and recognition is moving object extraction. After extracting it, perform representation model, or feature extraction prior to evaluate it with a classifier process. Failure in moving object extraction will contribute an error in detection and recognition process. This extraction is performed by processing of several frames to distinguish foreground from the background. In this paper a simple extraction method which was suitable for real time application is developed. The first stage, moving object detection by using frames subtraction is performed. This yielded boundary label in the motion area. This boundary label couldn't extract the whole moving object. For every row of image, calculate the boundary density function. The probability of pixels in this density is used to extract them as foreground. The experimental result, this system can extract the moving object in computation time of 71 - 80 ms, video rate of 15 frames per second (fps) and image frame size of 640x480 pixels, which is suitable for real time application. In this paper proposed method to extract the moving object is demonstrated. It can extract the object

fit the object shape, in shortly computation time. It is suitable for real time application. Due to this proposed method relies on the boundary pixels, there are some foreground pixels recognized as the background.

Paper 24: "Rich feature hierarchies for accurate object detection and semantic segmentation", by Ross Girshick, Jeff Donahue, Trevor Darrell, Jitendra Malik [24]

Description: Object detection performance, as measured on the canonical PASCAL VOC dataset, has plateaued in the last few years. The best-performing methods are complex ensemble systems that typically combine multiple low-level image features with high-level context. In this paper, a simple and scalable detection algorithm that improves mean average precision (mAP) by more than 30% relative to the previous best result on VOC 2012—achieving a mAP of 53.3% is proposed. The approach combines two key insights: (1) one can apply high-capacity convolutional neural networks (CNNs) to bottom-up region proposals in order to localize and segment objects and (2) when labeled training data is scarce, supervised pre-training for an auxiliary task, followed by domain-specific fine-tuning, yields a significant performance boost. Since it combine region proposals with CNNs, it is called R-CNN: Regions with CNN features In recent years, object detection performance had stagnated. The best performing systems were complex ensembles combining multiple low-level image features with high-level context from object detectors and scene classifiers. This paper presents a simple and scalable object detection algorithm that gives a 30% relative improvement over the best previous results on PASCAL VOC 2012. This performance is achieved through two insights. The first is to apply high-capacity convolutional neural networks to bottom-up region proposals in order to localize and segment objects. The second is a paradigm for training large CNNs when labeled training data is scarce. Lets show that it is highly effective to pre-train the network with supervision for a auxiliary task with abundant data (image classification) and then to fine-tune the network for the target task where data is scarce (detection). It conjecture that the "supervised pre-training/domain-specific finetuning" paradigm will be highly effective for a variety of data-scarce vision problems. Lets conclude by noting that it is significant that these results were achieved by using a combination of classical tools from computer vision and deep learning (bottomup region proposals and

convolutional neural networks). Rather than opposing lines of scientific inquiry, the two are natural and inevitable partners.

Paper 25: "Deep Neural Networks for Object Detection", by Christian Szegedy, Alexander Toshev, Dumitru Erhan [25]

Description: Deep Neural Networks (DNNs) have recently shown outstanding performance on image classification tasks. In this paper lets go one step further and address the problem of object detection using DNNs, that is not only classifying but also precisely localizing objects of various classes. Here present a simple and yet powerful formulation of object detection as a regression problem to object bounding box masks. Lets define a multi-scale inference procedure which is able to produce high-resolution object detections at a low cost by a few network applications. State-of-the-art performance of the approach is shown on Pascal VOC. In this work leverage the expressivity of DNNs for object detector. It show that the simple formulation of detection as DNN-base object mask regression can yield strong results when applied using a multi-scale course-to-fine procedure. These results come at some computational cost at training time – one needs to train a network per object type and mask type. As a future work lets aim at reducing the cost by using a single network to detect objects of different classes and thus expand to a larger number of classes.

Paper 26: "Fast, Accurate Detection of 100,000 Object Classes on a Single Machine", by Thomas L. Dean, Mark A. Ruzon, Mark E. Segal, Jonathon Shlens, Sudheendra Vijayanarasimhan, J. Yagnik [26]

Description: Many object detection systems are constrained by the time required to convolve a target image with a bank of filters that code for different aspects of an object's appearance, such as the presence of component parts. Locality-sensitive hashing to replace the dot-product kernel operator in the convolution with a fixed number of hash-table probes that effectively sample all of the filter responses in time independent of the size of the filter bank is exploited. To show the effectiveness of the technique, lets apply it to evaluate 100,000 deformable-part models requiring over a million (part) filters on multiple scales of a target image in less than 20 seconds using

a single multi-core processor with 20GB of RAM. This represents a speed-up of approximately 20,000 times - fthis orders of magnitude - when compared with performing the convolutions explicitly on the same hardware. While mean average precision over the full set of 100,000 object classes is around 0.16 due in large part to the challenges in gathering training data and collecting ground truth for so many classes and achieved a mAP of at least 0.20 on a third of the classes and 0.30 or better on about 20% of the classes.

Paper 27: "Performance Analysis and CPU vs GPU Comparison for Deep Learning", by Ebubekir BUBER, Banu DIRI [27]

Description: Deep learning approaches are machine learning methods used in many application fields today. Some core mathematical operations performed in deep learning are suitable to be parallelized. Parallel processing increases the operating speed. Graphical Processing Units (GPU) are used frequently for parallel processing. Parallelization capacities of GPUs are higher than CPUs, because GPUs have far more cores than Central Processing Units (CPUs). In this study, benchmarking tests were performed between CPU and GPU. Tesla k80 GPU and Intel Xeon Gold 6126 CPU was used during tests. A system for classifying Web pages with Recurrent Neural Network (RNN) architecture was used to compare performance during testing. CPUs and GPUs running on the cloud were used in the tests because the amount of hardware needed for the tests was high. During the tests, some hyperparameters were adjusted and the performance values were compared between CPU and GPU. It has been observed that the GPU runs faster than the CPU in all tests performed. In some cases, GPU is 4-5 times faster than CPU, according to the tests performed on GPU server and CPU server. These values can be further increased by using a GPU server with more features.

Paper 28 : "Image Quality Assessment through Contour Detection" by Shuo Liu, Xiang Peng and Zheng Liu [28]

Description :

It presenst study on objective image quality assessment from the segmentation perspective. Instead of modeling the image distortion or degradation process, the

assessment is achieved from the end user's point of view, i.e. how the image is interpreted by segmentation operation. The accessibility of informative features in an image subject to certain degradation is evaluated by comparing the contour of input and reference images. A new quality index is defined with the F-measure in the contour mapping measure framework. The demonstrative experiments illustrate how the index works for the prepared images and the preliminary results prove the effectiveness of the proposed approach. an image quality index based on image segmentation and contour mapping measure is proposed. It provide a different perspective for the image quality assessment in comparison with the well-known SSIM index. In this study, the gPb-ucm algorithm was adopted to obtain the edge strength map for the input image. The contour derived from the strength map is used in the precision-recall framework to define the image quality index. Preliminary results demonstrate the effectiveness of the new index. Extensive experiments will be carried out to further validate the proposed index with the major datasets currently available for image quality assessment research. Moreover, the edge strength map may be used in the similarity assessment directly in our future work. Currently, all indices use a single value to represent image quality. However, image quality assessment is actually a multi-attribute phenomenon. Although an overall quality index is convenient, it is still questionable to claim two images undergoing different degradation of the same quality. Thus, multi-channel or multi-attribute assessment is preferred while a combined/fused index could summarize all these channels or attributes. Future research may also consider identifying and assessing the degradation process in the image quality assessment.

Paper 29 : " Image Quality Assessment Using Multi-Method Fusion" by T.-J. Liu, W. Lin, and C.-C. Kuo [29]

Description :

A new methodology for objective image quality assessment (IQA) with multi-method fusion (MMF) is presented in this work. The research is motivated by the observation that there is no single metric that can give the best performance in all situations. To achieve MMF, we adopt a regression approach. The new MMF score is set to be the nonlinear combination of scores from multiple methods with suitable weights obtained by a training process. In order to improve the regression results further, we divide

distorted images into three to five groups based on the distortion types and perform regression within each group, which is called context-dependent MMF (CD-MMF). One task in CD-MMF is to determine the context automatically, which is achieved by a machine learning approach. To further reduce the complexity of MMF, we perform algorithms to select a small subset from the candidate method set. The result is very good even if only quality assessment methods are included in the fusion process. The proposed MMF method using support vector regression (SVR) is shown to outperform a large number of existing IQA methods by a significant margin when being tested in six representative databases. As far as we know, there is no single existing image quality index gives the best performance in all situations. Thus, in this work, we have proposed an open, inclusive framework for better performance with the current level of technology and for easy extension when new technology emerges. To be more specific, we have presented a multi-method fusion (MMF) approach for image quality assessment and proposed two MMF-based quality indices, based upon machine learning. It was shown by experiments with six different databases (totally images) that both of them outperform state-of-the-art quality indices by a significant margin. As expected, the complexity of the MMF method is higher since it involves the calculation of multiple methods. However, with the help of the algorithms (SFMS and BIRD), we can reduce the number of fused methods and lower the complexity of MMF. As long as we can keep the number of fused IQA methods not more than three, the computational time of MMF is around minute per image.

Paper 30 : "Image quality assessment: From error visibility to structural similarity" by Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli [30]

Description :

Earth is inhomogeneous, which means its elastic characteristics change with depth. The seismic method employs the propagation of waves throughout the earth to locate different structures and stratigraphy. Understanding the wave propagation is an important matter in exploration seismology; therefore modeling of seismic wave is an important tool. To validate the interpreted earth model out of the seismic data, seismic synthetic seismograms should be generated in a process named "seismic forward modeling". Finite difference method is used as one of the most common numerical

modeling techniques. In this paper the accuracy of finite difference method in seismic section modeling is explored on different modeled data set of heterogeneous earth. It is shown that finite difference method completes with migration to reposition the events in their correct location. Two different migration methods are used and various velocities are also tested to determine an appropriate migration velocity. bjective methods for assessing perceptual image quality traditionally attempted to quantify the visibility of errors (differences) between a distorted image and a reference image using a variety of known properties of the human visual system. Under the assumption that human visual perception is highly adapted for extracting structural information from a scene, we introduce an alternative complementary framework for quality assessment based on the degradation of structural information. As a specific example of this concept, we develop a Structural Similarity Index and demonstrate its promise through a set of intuitive examples, as well as comparison to both subjective ratings and state-of-the-art objective methods on a database of images compressed with JPEG and JPEG.

Paper 31 : "A universal image quality index" by Z. Wang and A. C. Bovik [31]

Description :

It presents an approach to surface image quality assessment for surface pattern and object recognition, classification, and identification has been described. The surface quality assessment finds many industrial applications such as auto-mated, advanced, and autonomous manufacturing processes. Given that in most industrial applications the target surface is an unknown variable, having a tool to measure the quality of the surface in real time has a significant value. To add to the complication, in most industrial applications, the surface (and therefore its image) suffers from several physical phenomena such as noise (of several different kinds), time, phase, and frequency shifts, and other clutter caused by interference and speckles. The proposed tool should also be able to measure the level of deterioration of the surface due to these environmental effects. Instead of using traditional error summation methods, the proposed index is designed by modeling any image distortion as a combination of three factors: loss of correlation, luminance distortion, and contrast distortion. Although the new index is mathematically defined and no human visual system model is explicitly employed, our experiments on various image distortion types indicate that it performs

significantly better than the widely used distortion metric mean squared error. Therefore, evaluation of quality of a surface is not an easy task. It requires a good understanding of the processing methods used and the types of environmental processes affecting the surface.

Paper 32: "Edge strength similarity for image quality assessment" by X. Zhang, X. Feng, W. Wang, and W. Xue

Description :

The objective image quality assessment aims to model the perceptual fidelity of semantic information between two images. In this letter, we assume that the semantic information of images is fully represented by edge-strength of each pixel and propose an edge-strength-similarity-based image quality metric (ESSIM). Through investigating the characteristics of the edge in images, we define the edge-strength to take both anisotropic regularity and irregularity of the edge into account. The proposed ESSIM is considerably simple, however, it can achieve slightly better performance than the state-of-the-art image quality metrics as evaluated on six subject-rated image databases.This letter presents a FR IQA metric based on theedge-strength similarity, which complementally exploits theanisotropic regularity and irregularity of the edges presentedin images. The experiment results on the subject-rated imagedatabases indicate that the proposed metric is quite competitivein terms of both assessment performance and computational complexity.

3. ANALYSIS

3.1 Detailed Statement of the problem

The problem of human detection in surveillance video involves developing a software system that can accurately detect and recognize humans in real-time and provide relevant information to the user. The system is designed to reduce the human effort required during investigations by automatically detecting and identifying individuals in surveillance footage. The system would require the integration of different technologies such as computer vision, machine learning, and data analytics to provide accurate detections and recognition.

The primary objective of the system would be to provide investigators with timely and accurate information on the whereabouts of individuals by detecting and recognizing them in surveillance footage. This would help investigators to identify and track suspects, as well as to locate missing persons or individuals of interest. The system would also help to reduce the workload of investigators by automatically analyzing and processing large amounts of surveillance footage, thereby saving time and resources.

To achieve these objectives, the system would need to be able to detect and recognize humans in real-time using computer vision and machine learning algorithms. The system would also need to be able to match the detected individuals with the ones present in the database, thereby providing relevant information about them to the user. The system would need to be designed to be user-friendly and accessible to investigators, even those who may not have a technical background. The system would need to provide simple and clear information that can be easily understood and used by investigators.

The system would require the integration of multiple technologies, including computer vision, machine learning, and data analytics. The computer vision technology is used to detect and recognize humans in surveillance footage. The machine learning algorithms are used to process this data and generate predictive models that can

accurately detect and recognize humans in real-time. The data analytics tools are used to analyze the data and generate insights that can help investigators make informed decisions about their investigations.

The system will require a user interface that is easy to use and accessible to investigators with varying levels of technical expertise. The interface will provide investigators with visualizations of the surveillance footage, including the timestamp and details of the detected individuals. The system will also need to be scalable and adaptable to different surveillance systems, environments, and lighting conditions.

The successful development of a human detection in surveillance video system can help investigators to more efficiently and accurately identify and track suspects, locate missing persons, and reduce the workload of investigations. This will result in more efficient and effective law enforcement, increased public safety, and improved quality of life for communities.

3.2 Requirement Specification

In this section we will look towards the Software and Hardware required for the implementation of the project. The software and hardware requirements for the implementation of the project will depend on the specific goals and needs of the system. Careful consideration and planning of these requirements are crucial to ensure the system's success and optimal performance. We have divided the requirements in two parts Software requirement and Hardware requirement.

3.2.1 Software Requirement

The project will require a programming language to implement the algorithms and logic for the system. The choice of programming language will depend on the specific needs and goals of the project. Additionally, the system may require a database management system to store and manage data such as user information, system logs, and other relevant data.

The project also require various software libraries and frameworks for image processing, computer vision, and machine learning. These libraries can be used to extract features from images, train and test machine learning models, and perform various image analysis tasks.

Software requirements for our project are described as follows:

- Python
- Django
- MySQL
- CV2
- PIL
- Face-recognition Library
- Numpy
- Pandas
- Dlib
- VS Code
- XAMPP

3.2.2 Hardware Requirement

In this project, a computer with sufficient processing power is needed. This project requires too much processing power, due to the image and video batch processing. The project may require a camera or multiple cameras to capture video footage for analysis. The camera specifications such as resolution, frame rate, and field of view will depend on the specific needs of the project. Additionally, the system may require a powerful processor, sufficient memory, and storage capacity to perform real-time analysis of the video stream.

Furthermore, the project may require additional hardware such as sensors, actuators, and other components to enable the system to interact with the environment. For instance, if the system is used for access control, it may require a sensor to detect when a person enters or leaves a specific area and an actuator to control access to that area.

Although in our project we haven't implemented sensors in future we can. Hardware requirements for our system are as follows:

- Laptop/PC
- System: Intel Processor i3/i5/i7 or AMD processors
- RAM
- Hard Disk: 20 GB
- Camera

3.3 Functional Requirement

Functional requirements are an essential aspect of software development. These requirements detail the features and functions that a software system must have to perform its intended tasks. They explain how the system should operate and what it needs to do to achieve its goals. A clear understanding of these requirements is crucial to ensuring that the software system is developed in a way that meets the expectations of its users and stakeholders.

To describe the functional requirements of a software system, it is common to use various techniques such as structured analysis or object-oriented analysis. Structured analysis is a method for identifying the data flows, processes, and data stores that are required to support a particular software system. This technique uses diagrams to depict the various elements of the system and how they interact with each other. These diagrams can be used to create a detailed blueprint for the system's functionality.

Object-oriented analysis, on the other hand, is a technique that focuses on identifying the objects and classes that are required to support the software system. This technique involves creating an analysis class diagram that depicts the classes, attributes, and methods required to support the system. Each class is described in detail, including its attributes and the operations it performs.

Regardless of the technique used, the functional requirements of a software system should cover all the major software functions required to achieve its goals. These

functions should be detailed in a way that provides a clear understanding of how the system will operate.

These functions can be further detailed using data flow diagrams or analysis class diagrams to provide a comprehensive understanding of how they interact with each other and how they contribute to the overall functionality of the system.

Overall functional requirements are a critical aspect of software development. They provide a detailed description of the features and functions required for the software system to accomplish its tasks. By using techniques such as structured analysis or object-oriented analysis, these requirements can be detailed in a way that provides a clear understanding of how the system will operate and how its various functions interact with each other.

3.3.1 Data Flow Diagram

A data flow diagram (DFD) is a graphical representation of the flow of data through a system. It provides a visual representation of the system's processes, inputs, outputs, and data stores. A DFD is a useful tool for analyzing, modeling, and documenting a system's functionality, as it allows developers to understand how data moves through a system and how it is transformed along the way.

DFDs are divided into levels, with each level providing increasing levels of detail about the system being modeled. The highest-level DFD provides a broad overview of the system, while subsequent levels provide more detailed views of the system's processes, data flows, and data stores.

DFDs are widely used in software development, systems engineering, and business process management. They can be used to identify inefficiencies in a system, to analyze the impact of changes to a system, and to communicate the design of a system to stakeholders. Overall, a well-designed DFD can help to ensure that a system meets the needs of its users and operates efficiently and effectively.

DFD level – 1

Here Figure 3.1 shows DFD level – 1 indicates the basic flow of data in the system. In this System Input is given equal importance as that for Output.

- Input: This component represents the input to the system, which is uploading the video that needs to be analyzed.
- System: This is the main component of the system, where the video is processed to detect human faces and match them with the ones present in the database. It also includes the functionality to add details of persons to be detected, as well as view the details of detected persons.
- Output: This component represents the output of the system, which includes details about the detected persons such as their timestamp and frame when they were first detected. It also includes information about whether the person detected is a match with the one in the database or not.

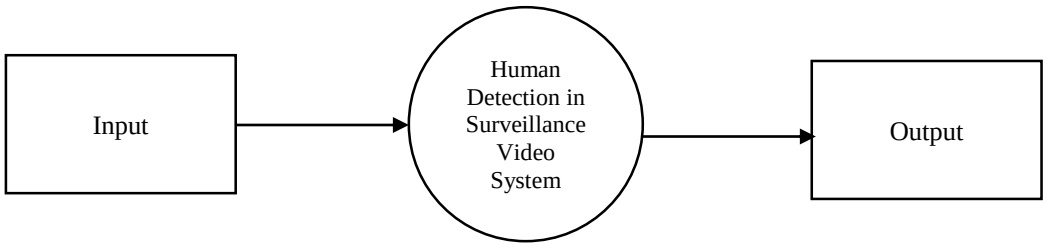


Figure. 3.1 Data Flow Diagram

3.4 Non-Functional Requirements

Non-functional requirements are the software specifications that describe the qualitative aspects of a software. It lists the desired qualitative features of a software or application, which don't fall under the category of any function/use-case. Non-functional features do not perform any action, instead they help in enhancing the software performance.

Accuracy:

- The system should have high accuracy in detecting humans, with a low false positive and false negative rate.
- The system should be able to detect humans in different lighting conditions, angles, and distances.

Performance:

- The system must be able to process video data in real-time to provide timely alerts and reduce response time in emergency situations.
- The system should be able to handle a large volume of video data and multiple requests simultaneously.

Security and Privacy:

- The system must ensure the security and privacy of the surveillance video data and personal information of individuals.
- The system should be compliant with data privacy regulations and have measures to prevent unauthorized access to the system and data.

Scalability and Adaptability:

- The system should be scalable and able to handle different camera types and configurations.
- The system should be adaptable to different environments, such as indoor and outdoor settings, and different weather conditions.

Reliability and Availability:

- The system should be reliable and available at all times to provide continuous surveillance.
- The system should have measures in place to prevent system failures and ensure quick recovery in case of failures.

Usability and User Experience:

- The system should have a user-friendly interface that is easy to navigate and understand.
- The system should provide clear and concise information to users, such as the location and details of detected humans.
- The system should provide alerts and notifications in a timely and effective manner.

Maintainability:

- The system should be designed in a modular and scalable way to facilitate maintenance and updates.
- The system should have clear and well-documented code to enable easy debugging and maintenance.
- The system should have a monitoring system to track errors, failures, and performance issues that may require maintenance.
- The system should have a backup and restore system to ensure data is not lost in the event of system failure or maintenance.

Cost-effectiveness:

- The system should be cost-effective to develop, deploy, and maintain to ensure its long-term viability and sustainability.
- The system should utilize open-source technologies and tools to minimize software licensing costs.
- The system should be designed to run efficiently on low-cost hardware to minimize hardware costs.
- The system should be designed to be easily upgradable and scalable to reduce the need for costly replacements or upgrades in the future.
- The system should have a clear and transparent pricing model to help customers understand the costs associated with using the system.

3.5 Feasibility Study

The aim of the feasibility study activity is to determine whether it would be the financially and technically feasible to develop the system or not. A feasibility studies is carried out from following different aspects:

Operational Feasibility

This assessment involves undertaking a study to analyze and determine whether and how well the organization's needs can be met by completing the project. Operational feasibility studies also examine how a project plan satisfies the requirements identified in the requirements analysis phase of system development. The system has been

developed for all the users who are interested in this product, irrespective of technical background. We have given a demo of our project to technical as well as non-technical users and all the users found the system user friendly.

Technical Feasibility

This assessment focuses on the technical resources available to the organization. It helps organizations determine whether the technical resources meet capacity and whether the technical team is capable of converting the ideas into working systems. Technical feasibility also involves the evaluation of the hardware, software, and other technical requirements of the proposed system.

Implementation Feasibility

This project can easily be made available online without much consideration of the hardware and software. The only required thing at the applicant's side is the Internet connection, which is a not difficult issue these days. After setting up the project, all the users can access and configure the system from any smartphone connected with the same network. Also, these particular modules can be controlled remotely through other devices.

Scheduling Feasibility

This assessment is the most important for project success; after all, a project will fail if not completed on time. In scheduling feasibility, an organization estimate how much time the project will take to complete. When these areas have all been examined, the feasibility analysis help identify any constraints the proposed project may face, including:

- Internal Project Constraints: Technical, Technology, Budget, Resource, etc.
- External Constraints: Logistics, Environment, Laws, and Regulations, etc.

3.6 Use Case Diagram

The purpose of a use case diagram is to capture the dynamic aspect of a system. Use case diagrams are used to gather the requirements of a system including internal and external influences. These requirements are mostly design requirements. Hence, when a system is analyzed to gather its functionalities, use cases are prepared and actors are identified as shown in figure 3.2.

Use case diagrams are an essential tool in software engineering that helps in the visual representation of a system's functionality. It provides a high-level view of the system's functionality and the actors involved in the system. It is a key part of the Unified Modeling Language (UML) and helps in gathering requirements for a system, including both internal and external factors.

The use case diagram consists of three main elements: actors, use cases, and the system boundary. Actors are external entities that interact with the system and can be either human or other systems. Use cases represent a specific functionality that the system provides to the actors. The system boundary represents the boundary of the system and separates the system from external entities.

An effective use case diagram can help your team discuss and represent:

- Scenarios in which your system or application interacts with people, organizations, or external systems.
- Goals that your system or application helps those entities (known as actors) achieve.
- The scope of your system.

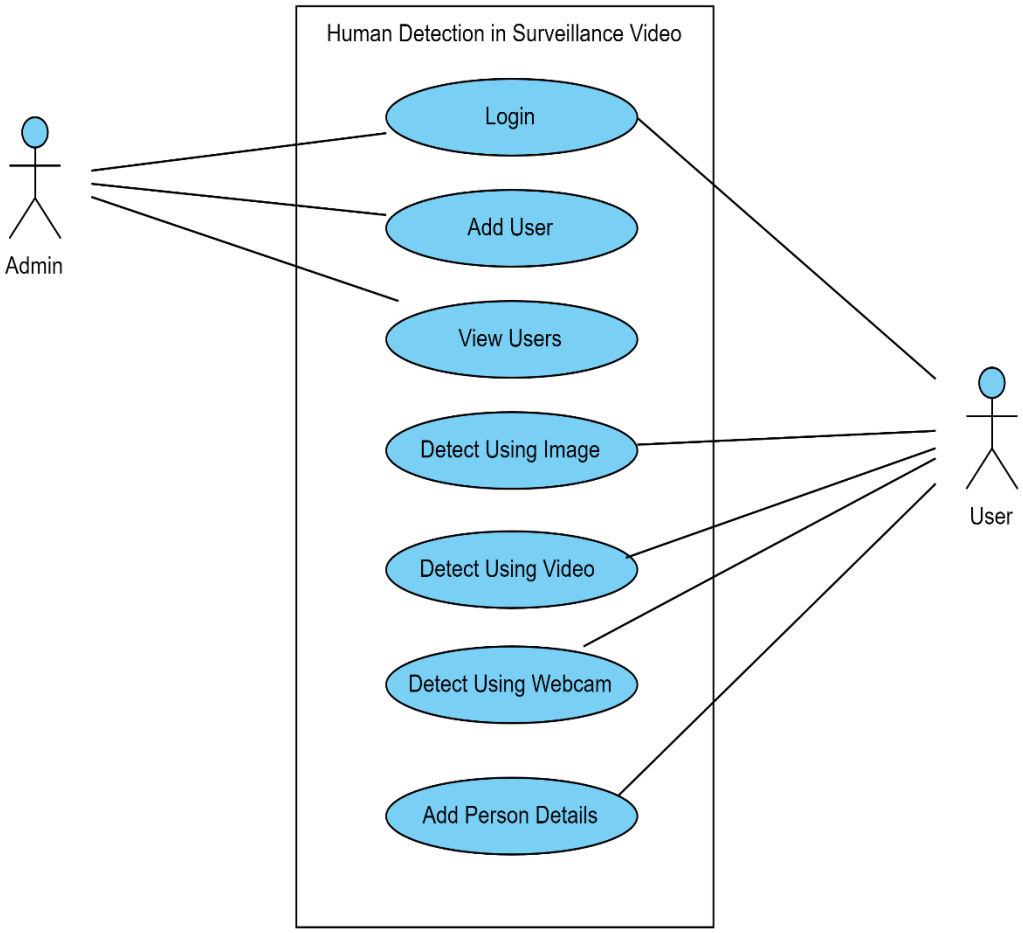


Figure 3.2 Use Case Diagram

3.7 Use Case Specification

Use case specification is a technique used in software engineering to capture and describe the functional requirements of a system from the end-user's perspective. It is a process of identifying and defining the interactions between the system and its users, as well as other systems or entities that the system interacts with. The goal of use case specification is to define the different scenarios or situations in which the system will be used, and to identify the interactions that take place between the system and its users or other external entities.

In the context of our human detection in surveillance video system, use case specification would involve identifying and defining the different scenarios or situations where the system would be used, and specifying the expected behavior of the system in each of these scenarios.

Use Case Name: Detect Wanted Persons in Surveillance Video
Actors: User, Admin, Surveillance Camera, Video Analytics System
Description: This use case involves detecting wanted persons in a surveillance video through the use of video analytics technology and providing timestamp and details of the person.
Preconditions:

- A surveillance camera is installed and operational.
- The video analytics system is up and running.
- The user has access to the system and the admin has granted the necessary permissions.

Basic Flow:

- User logs in to the system.
- User selects one of the three options on the dashboard to choose from: upload picture, upload video, or using webcam to model CCTV.
- User uploads or models the surveillance footage and provides additional information, such as location and time of the incident.
- Video analytics system analyzes the footage to identify the wanted person(s).

- If the wanted person(s) is detected, the system provides timestamp and details of the person, including their image and additional information.

- Admin can view the details of the detected persons and mark them as wanted, if necessary.
- User can download the timestamp and details of the detected person(s) for further investigation.

Alternate Flows:

A1. If the surveillance camera is not operational:

- The system does not receive any video footage and cannot detect wanted persons.
- No timestamp and details of the person(s) are provided.

A2. If the video analytics system is not operational:

- The surveillance camera captures video footage but it is not analyzed for detecting wanted persons.
- No timestamp and details of the person(s) are provided.

Postconditions:

- If wanted person(s) are detected, the system provides timestamp and details of the person(s).
- User and admin can view and download the details of the detected person(s) for further investigation.
- If no wanted person(s) are detected, no details are provided.

Exceptions:

- The system may generate false alarms if it detects persons who resemble the wanted person(s).
- The system may not detect the wanted person(s) in certain lighting or weather conditions.

4. DESIGN

4.1 Design Goal

The design goal of a Human Detection in Surveillance Video system is to reduce the amount of human effort required during investigations and provide quick and accurate identification of individuals in surveillance footage. To achieve this goal, the system would need to incorporate several key features and components, including:

- The system needs to collect and analyze video footage from surveillance cameras in real-time to detect human presence using computer vision algorithms.
- The system needs to use facial recognition technology to match detected faces with images in a database of known individuals and provide an alert when a match is found.
- Provide the exact frame when the person was first detected in the video, along with their name and other relevant details.
- Real-time processing of the video stream, using edge computing or cloud computing, to minimize latency and provide quick alerts to security personnel or law enforcement.
- Integration with existing surveillance systems, such as CCTV cameras, to provide a comprehensive view of the monitored area.
- Allow users to upload pictures of individuals they want to add to the database, and enable administrators to manage the database of known individuals.
- Provide an intuitive and user-friendly interface, such as a web-based dashboard, that allows users to easily search for and view surveillance footage, manage the database of known individuals, and adjust system settings.
- Completely open-source with good documentation to explain all the steps needed to build. For that, we aim to provide a web-based application that the user can use to control.

4.2 Design Strategy

As we have researched a lot about how and what to do in our project. As we go deeper and deeper into research, we come to know there are various aspects to do. That's why we figured out how we can go further and plan our task so that the requirements of our project get fulfilled. So here we go following the approach shown in Figure 4.1. We have divided into a certain tasks. They are as follows:

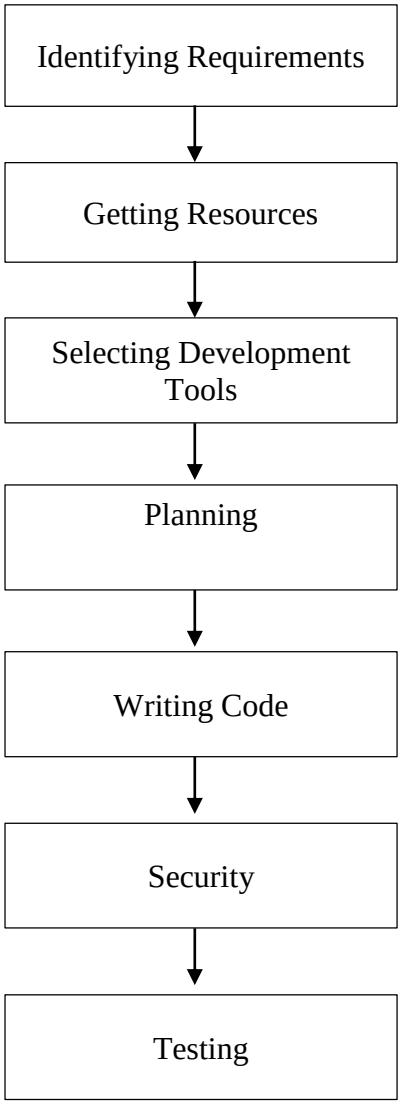


Figure 4.1 Strategy Diagram

Task 1: Identifying requirements

Firstly, we identified the requirements for this project. This task involves identifying the functionality required in the system. We analysed the company needs of what exactly the system they need to get designed.

Task 2: Gathering resources

We searched for best available resources for our project.

Task 3: Selecting Development Tools

In this step we selected the development tools that we found useful to build the interactive system. This involves choosing the programming language, backend framework, selecting database, etc.

Task 4: Planning

This step involves distributing task among group members.

Task 5: Writing Code

This step involves writing code for frontend, backend and database and developing our face recognition system.

Task 6: Adding security

Adding security by building a secured login system for user authentication.

Task 7: Testing

This step involves checking errors in the code, identifying the bugs, finding out whether the system provides the required functionalities, etc.

4.3 Architecture Diagram

An architecture diagram is a visual representation of the system's structure and organization, illustrating its components, their relationships, and how they interact to achieve the system's goals. Architecture diagrams help developers, designers, and stakeholders understand and communicate the design of the system.

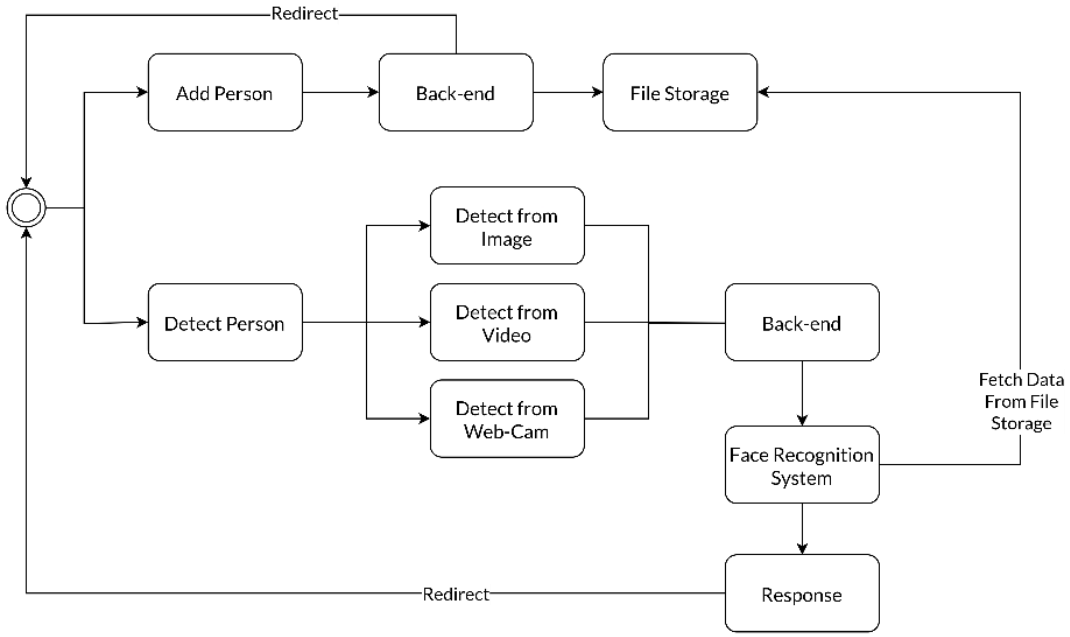


Figure 4.2 Architecture diagram

In this architecture diagram, the HDISV system is composed of several components that work together to detect wanted persons in a surveillance video. The surveillance camera here webcam, captures video footage, which is then transmitted to the face-recognition system. The face-recognition system analyzes the video footage to identify any person who is marked wanted present. If they are detected, the system an alert to the designated user or authorities. In this case it shows the timestamp of that persons first presence along with their details.

4.4 Class Diagram

Class diagram is a static diagram. It represents the static view of an application. Class diagram is not only used for visualizing, describing, and documenting different aspects of a system but also for constructing executable code of the software application.Class diagram describes the attributes and operations of a class and also the constraints imposed on the system. The class diagrams are widely used in the modeling of objectoriented systems because they are the only UML diagrams, which can be mapped directly with object-oriented languages.Class diagram shows a collection of classes, interfaces, associations, collaborations, and constraints. It is also known as a structural diagram.

Purpose of Class Diagrams

A class diagram is one of the most widely used UML diagrams for modeling the static view of an application. The purpose of a class diagram is to provide an overview of the classes and their relationships in a system. It is used to model the static structure of an object-oriented system, which includes the classes, their attributes, operations, and the relationships between them.

A class diagram is made up of classes, interfaces, and their relationships. The classes are represented as rectangles with their name, attributes, and methods listed inside. The attributes are the data members of the class and the methods are the operations that can be performed on the object. The relationships between classes are represented by lines connecting the classes.

Class diagrams are important in object-oriented programming as they provide a visual representation of the classes and their relationships in the system. They help to identify the key components of the system and how they relate to one another. This allows developers to better understand the system they are working on, and to make informed decisions about how to design and implement it.

The purpose of the class diagram can be summarized as −

- Analysis and design of the static view of an application.
- Describe responsibilities of a system.
- Base for component and deployment diagrams.
- Forward and reverse engineering.

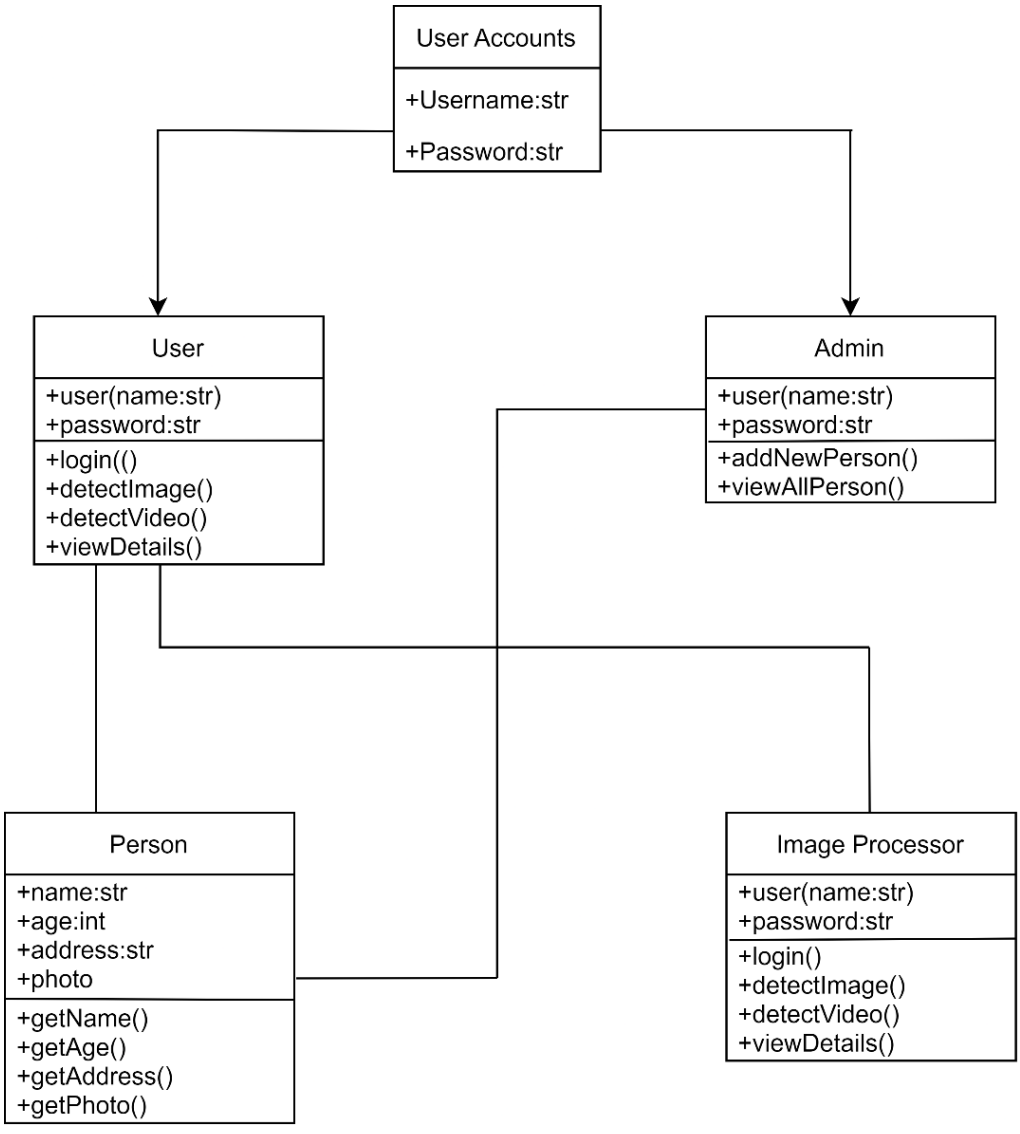


Figure 4.3 Class Diagram

4.4 Sequence Diagram

Sequence diagrams are a popular dynamic modeling solution in UML because the specifically focus on lifelines, or the processes and objects that live simultaneously, and the messages exchanged between them to perform a function before the lifeline ends. They are the most commonly used Interaction diagrams. The sequence diagram represents the flow of messages in the system and is also termed as an event diagram. It helps in envisioning several dynamic scenarios. It portrays the communication between any two lifelines as a time-ordered sequence of events, such that these lifelines took part in the run time.It conist of lifeline, actor and messags. An individual participant in the sequence diagram is represented by a lifeline. It is positioned at the top of the diagram. A role played by an entity that interacts with the subject is called as an actor. It is out of the scope of the system. It represents the role, which involves human users and external hardware or subjects. An actor may or may not represent a physical entity, but it purely depicts the role of an entity. Several distinct roles can be played by an actor or vice versa. The messages depict the interaction between the objects and are represented by arrows. They are in the sequential order on the lifeline. The core of the sequence diagram is formed by messages and lifelines.

Purpose of Sequence Diagrams

- To model high-level interaction among active objects within a system.
- To model interaction among objects inside a collaboration realizing a use case.
- It either models generic interactions or some certain instances of interaction

The figure 4.4, shows the sequence diagram for Human Detection in Surveillance Video system.

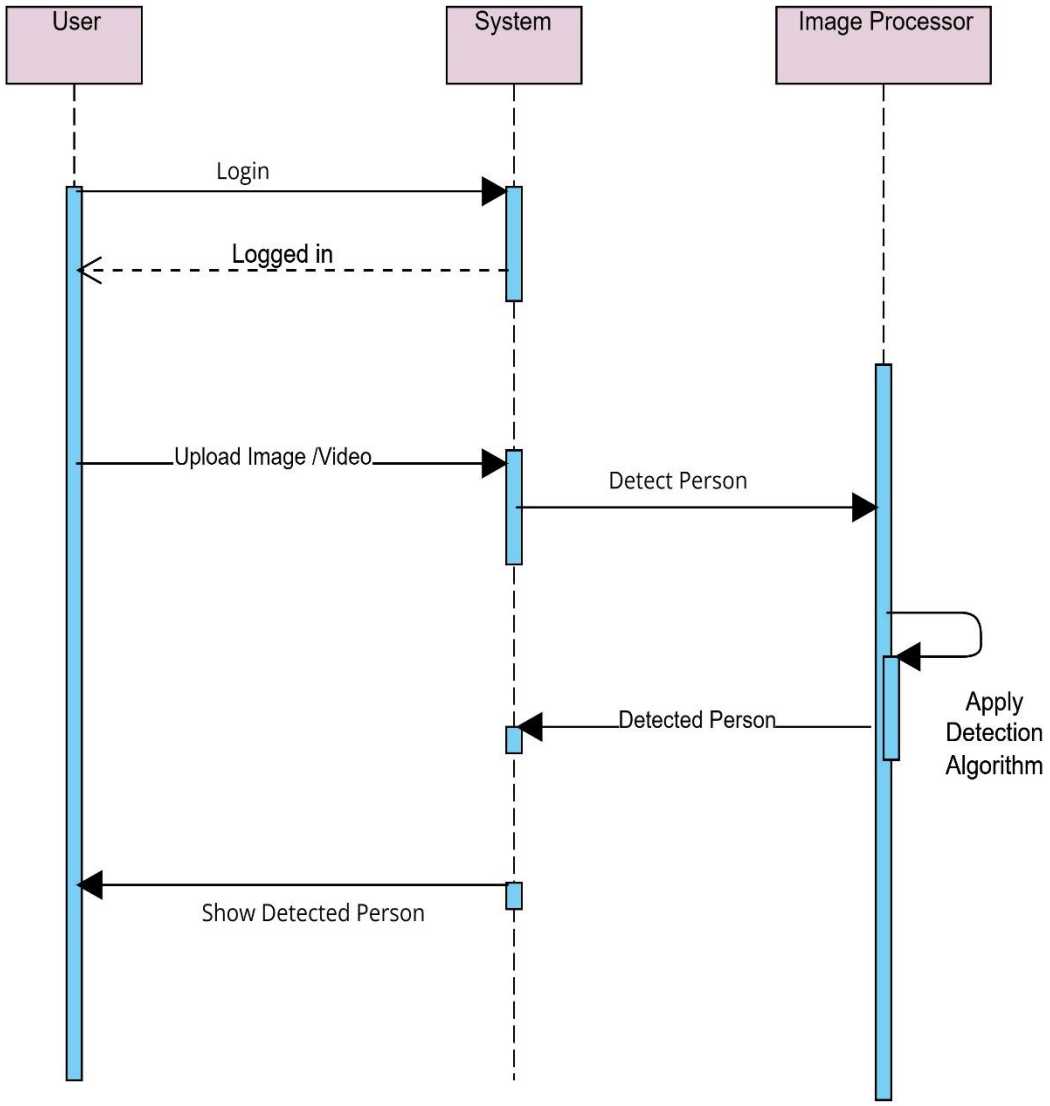


Figure 4.4 Sequence Diagram

4.5 Collaboration Diagram

The collaboration diagram is used to show the relationship between the objects in a system. Both the sequence and the collaboration diagrams represent the same information but differently. Instead of showing the flow of messages, it depicts the architecture of the object residing in the system as it is based on object-oriented programming. An object consists of several features. Multiple objects present in the system are connected to each other. The collaboration diagram, which is also known as a communication diagram, is used to portray the object's architecture in the system. The collaborations are used when it is essential to depict the relationship between the object. Both the sequence and collaboration diagrams represent the same information, but the way of portraying it is quite different. The collaboration diagrams are best suited for analyzing use cases.

Purpose of Collaboration Diagrams

A collaboration diagram, also known as a communication diagram, is a type of interaction diagram that shows how objects collaborate to accomplish a particular task or function within a system. It illustrates the dynamic interactions between objects and the messages exchanged between them to accomplish a particular task.

In a collaboration diagram, objects are represented as rectangular boxes, and the messages exchanged between them are represented as arrows. The arrows show the direction of the message flow between the objects, and may include information about the message content and timing.

The collaboration diagram is useful for showing the relationships between objects and the sequence of interactions required to accomplish a task or function. It can be used to model complex systems or processes, and can help developers identify potential issues or bottlenecks in a system's design.

Overall, the collaboration diagram is a powerful tool for visualizing the interactions

- The collaboration diagram is also known as Communication Diagram.
- It mainly puts emphasis on the structural aspect of an interaction diagram, i.e., how lifelines are connected.
- The syntax of a collaboration diagram is similar to the sequence diagram; just the difference is that the lifeline does not consist of tails.
- The messages transmitted over sequencing is represented by numbering each individual message.
- The collaboration diagram is semantically weak in comparison to the sequence diagram.
- The special case of a collaboration diagram is the object diagram.
- It focuses on the elements and not the message flow, like sequence diagrams.

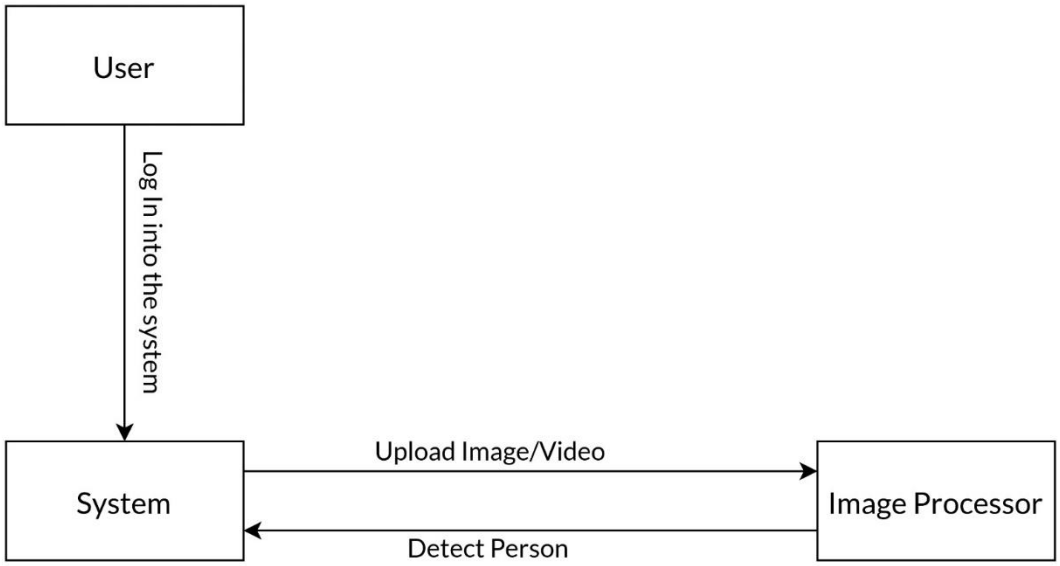


Figure 4.5 Collaboration Diagram

4.6 State Chart Diagram

State chart diagrams provide us an efficient way to model the interactions or communication that occurs within the external entities and a system. These diagrams are used to model the event-based system. A state of an object is controlled with the help of an event. State chart diagrams are used to describe various states of an entity within the application system State chart diagrams provide a visual representation of the behavior of a system. They can help developers and stakeholders understand the interactions between different components of the system and identify areas where improvements can be made. By modeling the behavior of a system in this way, it becomes easier to test and validate the system's functionality. State chart diagrams are an essential tool for software developers and engineers working on complex systems.

Purpose of State Chart Diagrams

State chart diagrams define different states of an object during its lifetime. A state is a condition in which an object exists and it changes when some event is triggered. The state chart diagram describes the flow of control from one state to another state. The transitions between states are triggered by events and are represented by arrows.
In a state chart diagram, an object is represented as a rectangle with its name inside. The states are represented by circles with their names inside. Transitions between states are represented by arrows, with the event that triggers the transition labeled on the arrow. Actions that are performed during a transition are shown in the form of activity diagrams.

The most important purpose of a state chart diagram is to model the lifetime of an object from creation to termination. This includes the different states that the object can be in and the events that cause it to transition from one state to another. By modeling the behavior of an object, it is possible to identify potential errors or issues that may arise during its lifetime.

State chart diagrams are also used for forward and reverse engineering of a system. Forward engineering involves using the state chart diagram to create the code for the

system, while reverse engineering involves using the code to create the state chart diagram. This can be useful for understanding the behavior of an existing system or for creating a new system from scratch.

Following are the main purposes of using State chart diagrams –

- To model the dynamic aspect of a system.
- To model the life time of a reactive system.
- To describe different states of an object during its life time.
- Define a state machine to model the states of an object

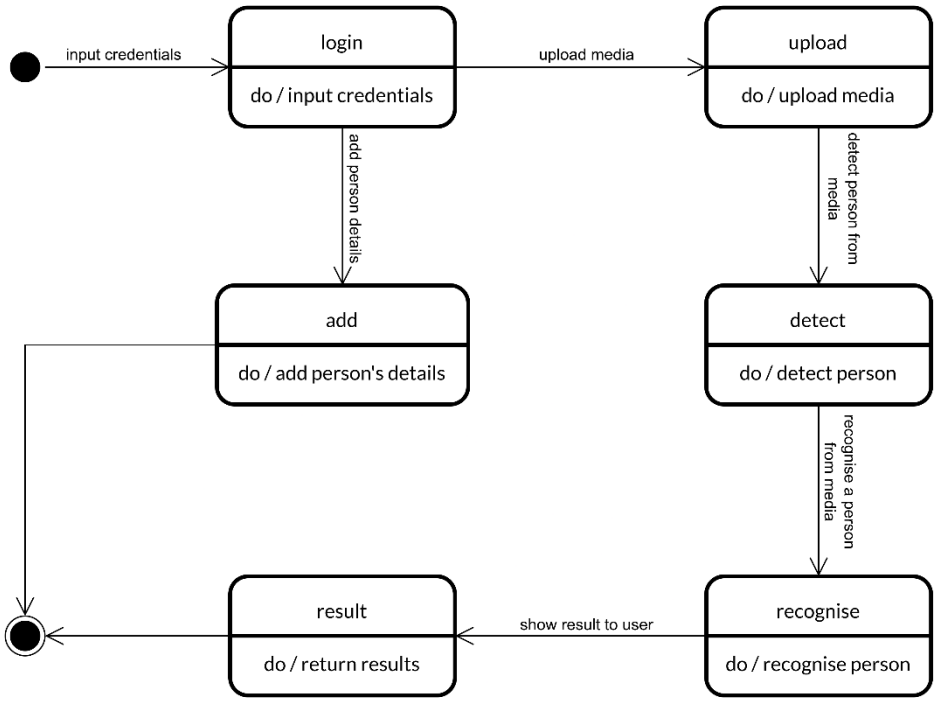


Figure 4.6 State Chart Diagram

4.7 Activity Diagram

The general work-flow of the planner can be graphically represented in an activity diagram. Figure 4.7 shows how user will use the system and the step-by-step process they will go through as they progress through the site. The diagram shows the workflow for all average user. The user is then able to interact with selected modules, or open new modules. Activity diagram is basically a flowchart to represent the flow from one activity to another activity. The activity can be described as an operation of the system. The control flow is drawn from one operation to another

Purpose of Activity Diagrams

The basic purpose of activity diagrams is similar to other UML diagrams. It captures the dynamic behavior of the system. Other UML diagrams are used to show the message flow from one object to another but the activity diagram is used to show message flow from one activity to another.

Activity is a particular operation of the system. Activity diagrams are not only used for visualizing the dynamic nature of a system, but they are also used to construct the executable system by using forward and reverse engineering techniques. The only missing thing in the activity diagram is the message part.

It does not show any message flow from one activity to another. Activity diagram is sometimes considered as the flowchart. Although the diagrams look like a flowchart, they are not. It shows different flows such as parallel, branched, concurrent, and single The purpose of an activity diagram can be described as:

- Draw the activity flow of a system.
- Describe the sequence from one activity to another.
- Describe the parallel, branched and concurrent flow of the system

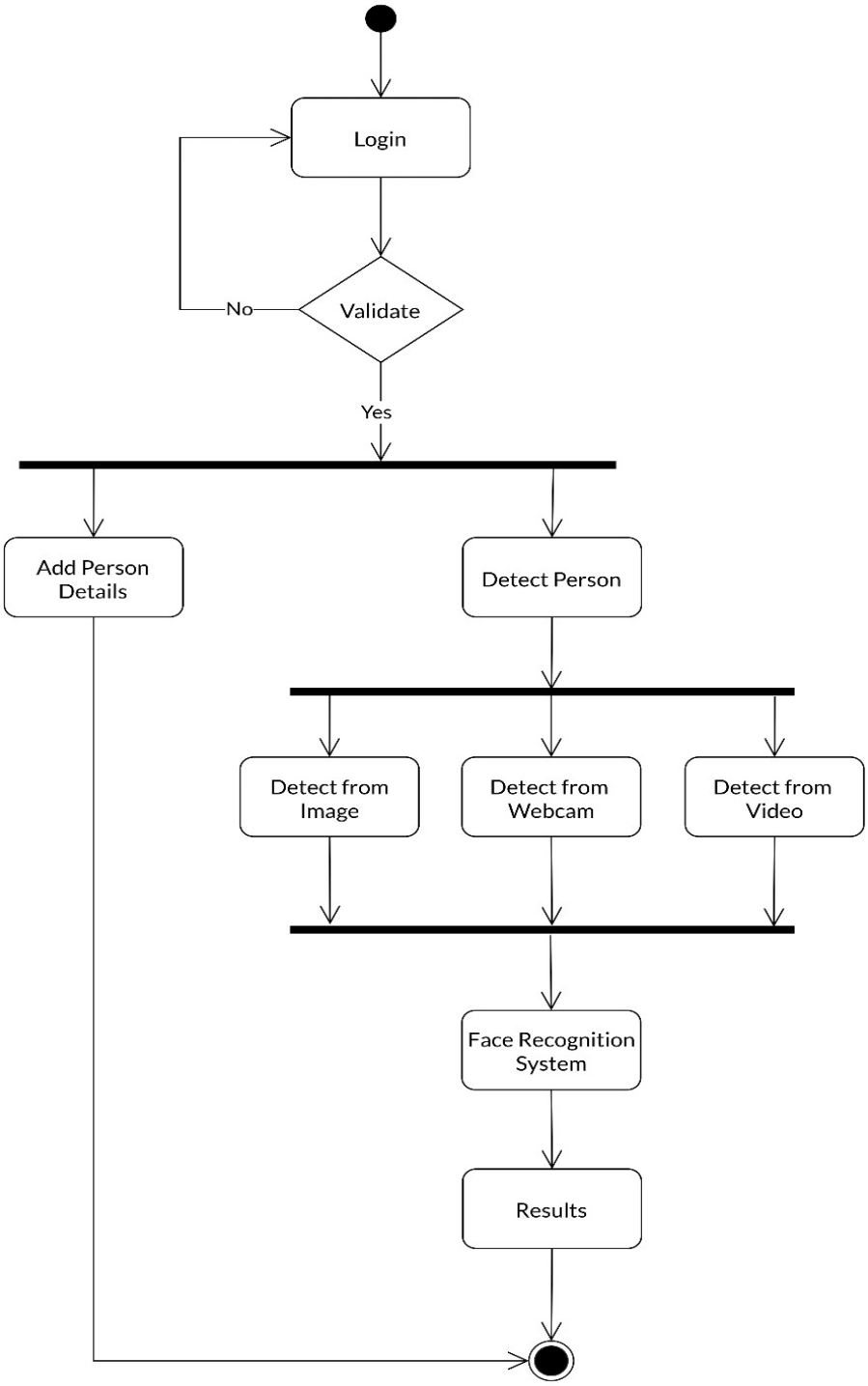


Figure 4.7 Activity Diagram

5. IMPLEMENTATION

5.1 Implementation Strategy

To implement the Human Detection in Surveillance Video system, we will follow a systematic approach. Firstly, we will research and gather information about different surveillance technologies and algorithms, including object detection and tracking, facial recognition, and machine learning. We will also analyze the specific requirements of law enforcement agencies and other potential users of the system. Based on our research and analysis, we will develop a detailed system architecture and design that meets the identified requirements and incorporates the latest technologies and algorithms.

Next, we will develop and test the system's core functionality, including video input and output, real-time object detection and tracking, and integration with other databases and surveillance systems. We will use a combination of open-source and proprietary software tools and libraries to build the system's components and integrate them into a cohesive whole. We will also conduct extensive testing and evaluation to ensure that the system meets its performance, accuracy, and reliability requirements.

We will develop a web interface and mobile application that allows law enforcement officers to access the system's functionality while in the field. The interface will provide real-time video feeds, object and person tracking, and other relevant information to help officers identify potential suspects and track their movements. We will also incorporate features such as push notifications, alerts, and remote access to enhance the system's usability and effectiveness.

We will continuously monitor and improve the system's performance and features based on user feedback and emerging technologies and algorithms. We will also work closely with law enforcement agencies and other users to identify and address any issues or concerns related to privacy, data security, and other legal and ethical considerations. The implementation of the Human Detection in Surveillance Video system can provide significant benefits to law enforcement agencies and improve public

safety by enabling more efficient and effective detection and tracking of potential suspects.

5.2 Hardware Platform Used

The hardware specifications for a human detection in surveillance video system will depend on various factors such as the scale of the system, the number of cameras, the complexity of algorithms used for detection, and the desired level of accuracy. Generally, a system that requires high-performance computing and real-time processing capabilities will need a powerful hardware configuration.

- Laptop/PC
- System: Intel Processor i3/i5/i7 or AMD processors
- RAM : 8 GB
- Hard Disk: 20 GB
- Camera

The hardware specifications for a human detection in surveillance video system will depend on the specific requirements of the system, but a high-performance CPU, a significant amount of RAM, high-resolution cameras, and storage devices with large capacity are essential. Dedicated GPUs can also greatly improve system performance, and various peripherals such as network adapters and UPS units may also be required. It is important to note that the hardware specifications mentioned above are only a general guideline, and the actual requirements may vary depending on the scale and complexity of the application. Additionally, hardware requirements may change as the application evolves and new features are added. Therefore, it is recommended to regularly assess and upgrade the hardware to ensure optimal performance of the application.

5.3 Software Platform Used

In software requirements, the project will require a programming language to implement the algorithms and logic for the system. The choice of programming

language will depend on the specific needs and goals of the project. Additionally, the system may require a database management system to store and manage data such as user information, system logs, and other relevant data. The project also requires various software libraries and frameworks for image processing, computer vision, and machine learning. These libraries can be used to extract features from images, train and test machine learning models, and perform various image analysis tasks.

Python

Python is a popular programming language used in web development for building web applications, web APIs, and dynamic websites. Here are some key areas where Python is used in web development:

- Web frameworks: Python has several web frameworks that simplify the process of building web applications. Django is a popular web framework that follows the Model-View-Controller (MVC) architecture and provides a lot of out-of-the-box features such as authentication, admin interface, and ORM. Flask is a lightweight framework that follows the Model-View-Template (MVT) architecture and allows developers to have more control over the application's structure. Pyramid is a flexible framework that supports various development styles and is suitable for building large-scale applications.
- Web scraping: Python is commonly used for web scraping, which involves extracting data from websites. There are several Python libraries like Beautiful Soup, Scrapy, and Requests that make web scraping easy and efficient.
- Web APIs: Python is often used to create web APIs (Application Programming Interfaces) that allow different applications to communicate with each other. Flask and Django Rest Framework are popular libraries for building web APIs.
- Server-side scripting: Python is used as a server-side scripting language, which means it is executed on the server-side to generate dynamic content that is sent to the client-side. Python can be used in conjunction with web servers like Apache and Nginx to create dynamic web pages.
- Machine learning: Python has a strong ecosystem for machine learning and data science, which can be used to build web applications that use artificial

intelligence and predictive models. Flask and Django are often used in conjunction with machine learning libraries like TensorFlow and Keras.

Python is important in development for several reasons:

- Versatility: Python is a versatile language that can be used for various web development tasks, including web scraping, web development, web APIs, server-side scripting, and machine learning.
- Large ecosystem: Python has a large ecosystem of libraries and frameworks that make it easy to develop web applications. There are several web frameworks like Django, Flask, Pyramid, and Bottle, as well as libraries like Beautiful Soup, Scrapy, Requests, and others.
- Easy to learn: Python is an easy-to-learn language that has a simple and readable syntax. This makes it easy for developers to understand and maintain the codebase, and to onboard new developers to the project.
- Rapid prototyping: Python's simplicity and ease-of-use make it ideal for rapid prototyping. Developers can quickly build and test web applications, and iterate on them based on feedback.
- Community support: Python has a large and active community of developers who contribute to the language and its ecosystem. This community provides support, documentation, and resources to help developers build better web applications.

Django

Django is a Python web framework that is widely used to build web applications. It provides a lot of features and tools that make it easy for developers to create powerful and complex web applications. Django is known for its "batteries-included" approach, which means that it comes with a lot of built-in features and functionality that make web development faster and easier.

Here are some reasons why developers choose Django for web development:

- Rapid development: Django provides a lot of built-in features and functionality that make it easy to build complex web applications quickly. This can be particularly useful for startups and other organizations that need to quickly test new ideas and iterate on them.
- Scalability: Django is a scalable web framework that can be used to build applications of any size. It provides tools for managing large databases, handling high traffic loads, and more.
- Security: Django is designed with security in mind. It provides built-in tools for protecting against common web application security vulnerabilities like SQL injection and cross-site scripting (XSS).
- Highly customizable: Django is a highly customizable web framework that allows developers to build web applications the way they want. It provides a lot of flexibility and can be easily extended with plugins and extensions.
- Excellent documentation: Django has excellent documentation that makes it easy for developers to learn and use the framework. The documentation covers everything from basic concepts to advanced features, and is regularly updated to keep up with changes in the framework.
- Active community: Django has a large and active community of developers who contribute to the framework and its ecosystem. This community provides support, documentation, and resources to help developers build better web applications.
- Good for large projects: Django is an excellent choice for building large web development projects, as it provides tools for managing large codebases and databases.
- Admin interface: Django provides a built-in admin interface that allows developers to manage their web applications without writing a lot of custom code.
- Support for multiple databases: Django supports multiple databases, including MySQL, PostgreSQL, Oracle, and SQLite.
- URL routing: Django provides a powerful URL routing system that makes it easy to map URLs to views.

MySQL

MySQL is an open-source relational database management system (RDBMS) that is widely used in web development. It is a popular choice for web developers because it is fast, reliable, and easy to use. Here are some key features of MySQL:

- Relational database: MySQL is a relational database, which means that it stores data in tables that are related to each other. This allows developers to easily organize and manage data in a structured way.
- ACID-compliant: MySQL is ACID-compliant, which means that it ensures that transactions are processed reliably and consistently. This ensures data integrity and prevents data corruption or loss.
- Fast and scalable: MySQL is designed to be fast and scalable, making it an excellent choice for web applications that require high performance and can handle a large amount of data.
- Open-source: MySQL is an open-source software, which means that it is freely available and can be modified and distributed by anyone. This makes it a popular choice for developers who want to use a reliable and flexible database system without the cost of licensing fees.
- Cross-platform compatibility: MySQL is compatible with a wide range of operating systems, including Windows, Linux, and macOS. This makes it a versatile option for web developers who need to work with different systems.
- Easy to use: MySQL is easy to learn and use, even for developers who are new to database management systems. It has a simple and intuitive user interface, and its command line interface is also straightforward.
- High availability: MySQL supports high availability through features such as replication and clustering. This allows developers to ensure that their web applications remain available and responsive even in the event of hardware or software failures.
- Security: MySQL has a range of security features, including user authentication, encryption, and auditing. This helps developers to protect sensitive data and prevent unauthorized access.

Visual Studio Code

Visual Studio Code is a free and open-source code editor developed by Microsoft for Windows, macOS, and Linux operating systems. It provides a rich and customizable user interface, along with a wide range of features designed to enhance productivity and make coding easier and more efficient.

- Support for multiple programming languages and frameworks, with built-in support for popular languages such as Python, JavaScript, TypeScript, and C++
- Powerful debugging and testing system with support for breakpoints, variable inspection, and step-by-step debugging, as well as integration with popular testing frameworks such as Jest and Mocha
- Integrated source control system with support for Git out-of-the-box, and integration with popular source control platforms such as GitHub, Bitbucket, and Azure DevOps
- Highly customizable with a range of themes and icons, and allows users to configure keyboard shortcuts and other preferences to suit their workflow
- Powerful extension API, which allows developers to create their own custom extensions to enhance the editor's functionality

XAMPP

XAMPP is a popular, open-source software package that provides a complete web development environment. It is designed to work on Windows, Linux, and macOS operating systems, and includes all the components required to run a web server, including Apache HTTP Server, MySQL, and PHP.

- XAMPP enables quick and easy setup of a local web server for testing and development purposes.
- Allows developers to simulate a live server environment on their own computer, making it easier to test and debug their web applications.
- Includes essential components for web development, such as Apache HTTP Server for serving web pages, MySQL for flexible database management, and PHP for creating dynamic and interactive web pages.

- Includes additional tools and utilities like phpMyAdmin for managing MySQL databases and FileZilla FTP Server for transferring files to and from the server.
- Provides an intuitive control panel for managing the web server and other components, including starting and stopping the server, configuring settings, and accessing logs and other useful information.
- Easy to install with just a few clicks, making it accessible even to those with little technical experience.

The software and hardware requirements for the implementation of the project will depend on goals and needs of the system. Careful consideration and planning of these requirements are crucial to ensure the system's success and optimal performance.

5.4 Deployment Diagram

Deployment diagrams are used to visualize the topology of the physical components of a system, where the software components are deployed.Deployment diagrams are used to describe the static deployment view of a system. Deployment diagrams consist of nodes and their relationships.

Deployment diagrams are mainly used by system engineers. These diagrams are used to describe the physical components (hardware), their distribution, and association.
Deployment diagrams can be visualized as the hardware components/nodes on which the software components reside.Software applications are developed to model complex business processes. Efficient software applications are not sufficient to meet the business requirements. Business requirements can be described as the need to support the increasing number of users, quick response time, etc.

To meet these types of requirements, hardware components should be designed efficiently and in a cost-effective way.Now-a-days software applications are very complex in nature. Software applications can be standalone, web-based, distributed, mainframe-based and many more. Hence, it is very important to design the hardware components efficiently.

Purpose of Deployment Diagrams

The term Deployment itself describes the purpose of the diagram. Deployment diagrams are used for describing the hardware components, where software components are deployed. Component diagrams and deployment diagrams are closely related.

Component diagrams are used to describe the components and deployment diagrams shows how they are deployed in hardware.UML is mainly designed to focus on the software artifacts of a system. However, these two diagrams are special diagrams used to focus on software and hardware components.

Most of the UML diagrams are used to handle logical components but deployment diagrams are made to focus on the hardware topology of a system. Deployment diagrams are used by the system engineers.

The purpose of deployment diagrams can be described as −

- Visualize the hardware topology of a system.
- Describe the hardware components used to deploy software components.
- Describe the runtime processing nodes.

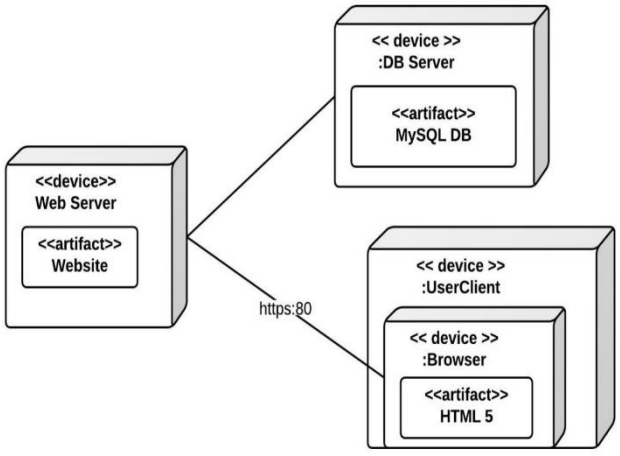


Figure 5.1 Deployment Diagram

5.5 Implementation Level Details

The implementation of the Human Detection in Surveillance Video system is a complex task that requires a expertise in computer vision and software development. The accuracy, scalability, and security of the system must be carefully considered to ensure its effectiveness in reducing human effort during investigations. Computer vision algorithms are used to analyze the video footage captured by surveillance cameras and identify human beings. We are using Face-recognition library which works exceptionally well for variable input.

- Preprocessing: Before feeding the surveillance video into the system, preprocessing may be required to remove any noise, enhance the quality, and adjust the color balance to improve the accuracy of human detection.
- Face-recognition: The human detection system will use a face-recognition library, to detect and identify human beings in the surveillance video. Further explanation is given in upcoming chapters.
- User Interface: The system should have a user-friendly interface that allows users to upload images or videos of wanted persons, add details about the person, and view the detected humans' details and timestamps.
- Integration with Surveillance System: The system needs to be integrated with the surveillance system to receive live video footage. Integration can be done through APIs, RTSP, or other standard interfaces. Although in our case we are using webcam to emulate CCTV.
- Alert System: When a wanted person is detected in the surveillance video, the system should generate an alert to the designated user or authorities through SMS, email, or push notifications or message on screen.
- Database Management: The system will need to manage a database of detected humans and wanted persons. The database should be secure and scalable, allowing the addition and deletion of data as required.

5.5.1 Dataset Details

In our project we are using Face-recognition libray which uses Labeled Faces in the Wild (LFW) is a database of face photographs designed for studying the problem of unconstrained face recognition. This database was created and maintained by researchers at the University of Massachusetts, Amherst (specific references are in Acknowledgments section). 13,233 images of 5,749 people were detected and centered by the Viola Jones face detector and collected from the web. 1,680 of the people pictured have two or more distinct photos in the dataset. The original database contains four different sets of LFW images and also three different types of "aligned" images. According to the researchers, deep-funneled images produced superior results for most face verification algorithms compared to the other image types. Hence, the dataset uploaded here is the deep-funneled version.

The LFW dataset is often used for training and testing face recognition algorithms. It is particularly challenging because of the large variations in poses, expressions, and lighting conditions. Many state-of-the-art face recognition algorithms have been trained and tested on this dataset, and it is considered to be a benchmark dataset in the field of face recognition.

The LFW dataset is freely available for research purposes, and can be downloaded from the official website of the dataset. Along with the images and labels, the dataset also includes bounding box annotations for the faces in each image, which can be used to train face detection algorithms.

To provide a standardized evaluation of face recognition algorithms on the LFW dataset, the LFW benchmark was created. The LFW benchmark is a widely used evaluation protocol for face recognition algorithms that involves dividing the dataset into two sets: a training set and a test set.

Researchers can submit their face recognition algorithms to the LFW benchmark website, where they will be evaluated on the test set according to several performance

metrics, including accuracy, true positive rate, false positive rate, and receiver operating characteristic curve.

The LFW dataset and benchmark have been instrumental in advancing the field of face recognition, and have spurred the development of many state-of-the-art algorithms and techniques. However, there is still much work to be done to improve the accuracy and robustness of face recognition algorithms, especially in challenging scenarios with low resolution, occlusion, and variability in appearance.

5.5.2 Methodology

We make use of the Face-recognition Library which is an open-source library for face recognition and detection tasks. It is written in Python and is built on the dlib-library, a contemporary C++ toolkit that includes machine learning methods and resources.

The working of the Face-recognition Library can be divided into three main parts: face-detection, face-alignment, and face-recognition.

Face Detection:

The crucial step in the Face-recognition Library is the face detection. The library uses an object detection algorithm to identify the location of faces in the input image. The library supports two different methods of face detection: the HOG-based method and the CNN-based method.

The HOG-based method uses a Histogram of Oriented Gradients (HOG) feature descriptor and a linear support vector machine (SVM) classifier to identify the presence of faces in an image. This method is faster but less accurate than the CNN-based method.

To find faces in an image, we'll start by making our image black and white because we don't need color data to find faces:

Figure 5.2 Black and White image of Person

Then we'll look at every single pixel in our image one at a time. For every single pixel, we want to look at the pixels that directly surrounding it:

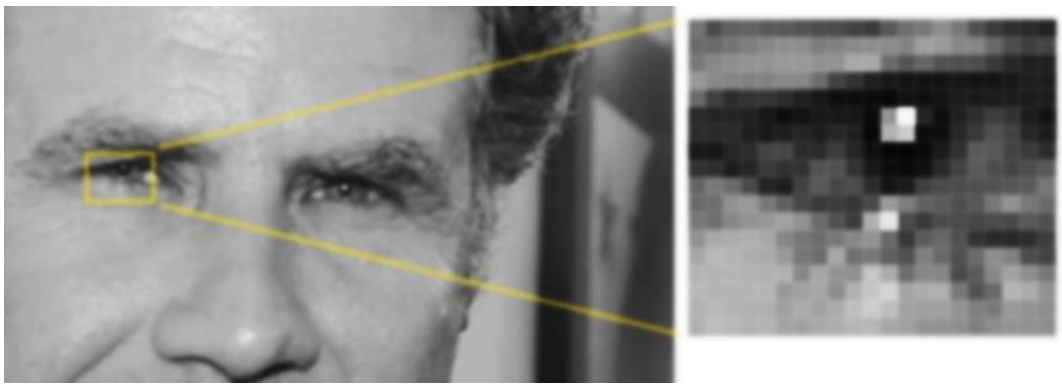


Figure 5.3 Pixel level representation of image

The objective is to determine the brightness level of the current pixel in comparison to the pixels that are immediately surrounding it. After that, an arrow will be drawn to indicate the direction in which the image is becoming darker.

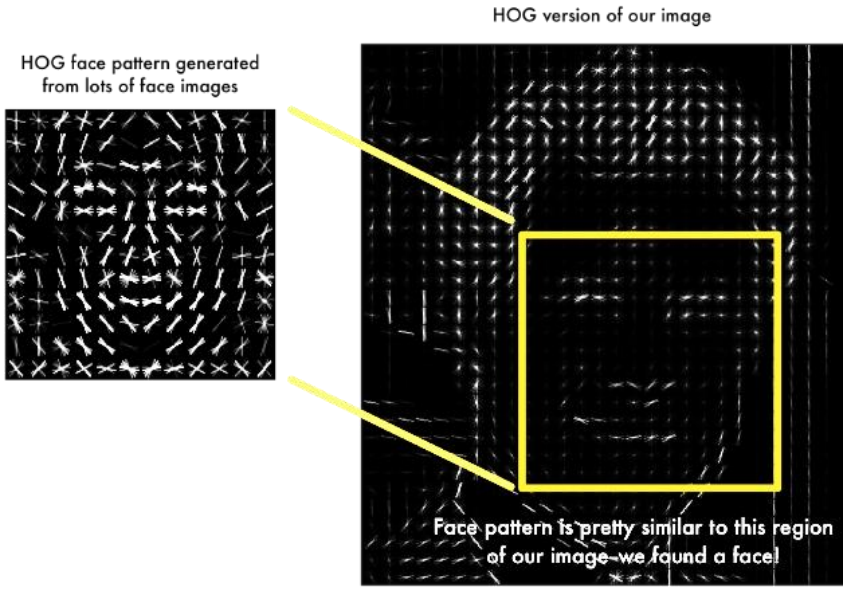


Figure 5.4 HOG version of face

The CNN-based method uses a deep convolutional neural network to detect faces in an image. This method is more accurate but slower than the HOG-based method.

Face Alignment:

Once the faces have been detected, the library uses face alignment to normalize the faces to a common coordinate system. The face alignment process involves identifying the facial landmarks, like the eyes, nose, and mouth, and then applying a geometric transformation to warp the image to a canonical face representation. This step helps to remove the variations in pose, scale, and orientation of the face images, making them more suitable for face recognition.

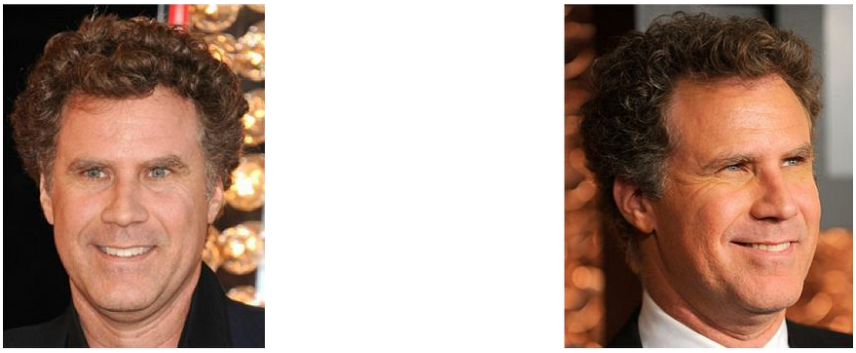


Figure 5.5 Views from different angle of same person

Humans can easily recognize that both images are of same person , but computers would see these pictures as two completely different people. To account for this, we will try to warp each picture so that the eyes and lips are always in the sample place in the image. This will make it a lot easier for us to compare faces in the next steps.

To do this, we are going to use an algorithm called face landmark estimation. There are lots of ways to do this, but we are going to use the approach invented in 2014 by Vahid Kazemi and Josephine Sullivan [33]. The basic idea is we will come up with specific points (called *landmarks*) that exist on every face — the top of the chin, the outside edge of each eye, the inner edge of each eyebrow, etc.:

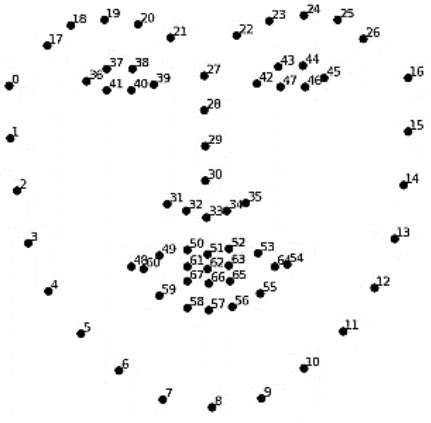


Figure 5.6 68 landmarks

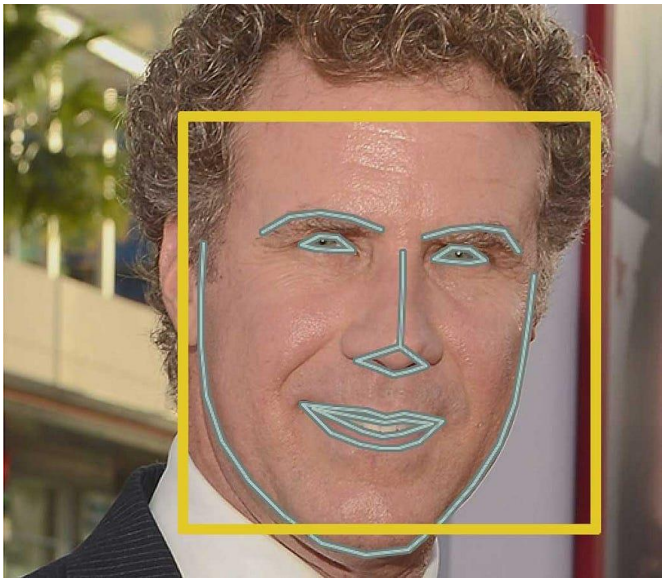


Figure 5.7 After mapping 68 landmarks on persons face

Now that the positions of the eyes and mouth have been identified, the image will be rotated, scaled, and sheared to center the eyes and mouth as accurately as possible. Only basic image transformations, such as rotation and scale, that preserve parallel lines (referred to as affine transformations) will be utilized, and there will be no advanced 3D warping techniques as they may distort the image.

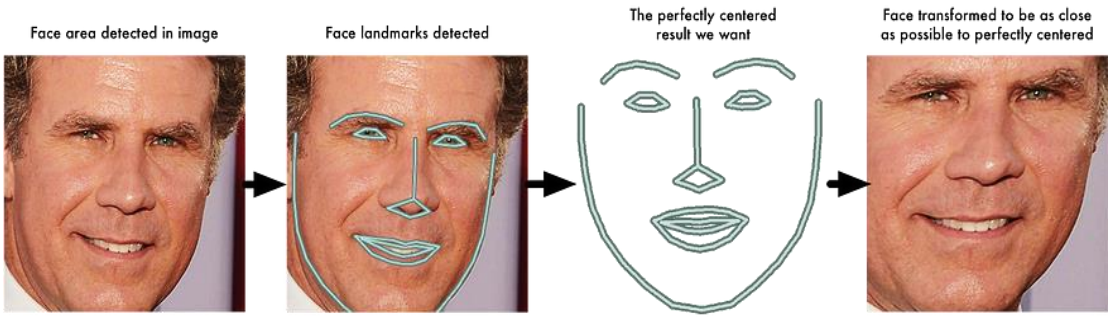


Figure 5.8 After affine transformations

Regardless of how the face is oriented, the eyes and mouth can now be centered in approximately the same location in the image. This will significantly enhance the accuracy of the subsequent step.

Face Recognition:

The final step in the Face-recognition Library is face recognition. The library uses a deep neural network to extract features from the aligned face images. These features are then compared to the database of known faces to identify the individual in the image. The library supports several different face recognition algorithms, including the classic Eigenfaces method, Fisherfaces method, and Local Binary Patterns Histograms (LBPH) method.

The core of the face recognition is process of encoding faces. The simplest approach is to compare the unknown face found in Step 2 with all the labeled pictures of people we have. However, this method is not feasible. Therefore, a new approach is needed, which involves extracting a few basic measurements from each face. This way, we can measure the unknown face in the same way and find the known face with the closest measurements. For instance, we may measure the size of each ear, the spacing between the eyes, the length of the nose, and other facial features.

So all we need to do ourselves is run our face images through their pre-trained network to get the 128 measurements for each face. The measurements for our test image is shown in figure 5.9.

The process of training a convolutional neural network to output face embeddings is computationally intensive and requires a significant amount of data. Even with an expensive NVidia Telsa video card, it can take up to 24 hours of continuous training to achieve high accuracy. However, once the network is trained, it can generate measurements for any face, even for those it has never seen before. Fortunately, the team at OpenFace has already published several pre-trained networks that can be used directly, saving the trouble of training a new network. By running the face images through their pre-trained network, we can obtain the 128 measurements for each face.

While it is unknown what specific parts of the face these 128 numbers measure, it is not relevant as long as the network generates nearly the same numbers when comparing two different pictures of the same person. If interested in trying this step, OpenFace provides a lua script that can generate embeddings for all images in a folder and write them to a csv file. We don't have to run it as the library we are using does this for us.

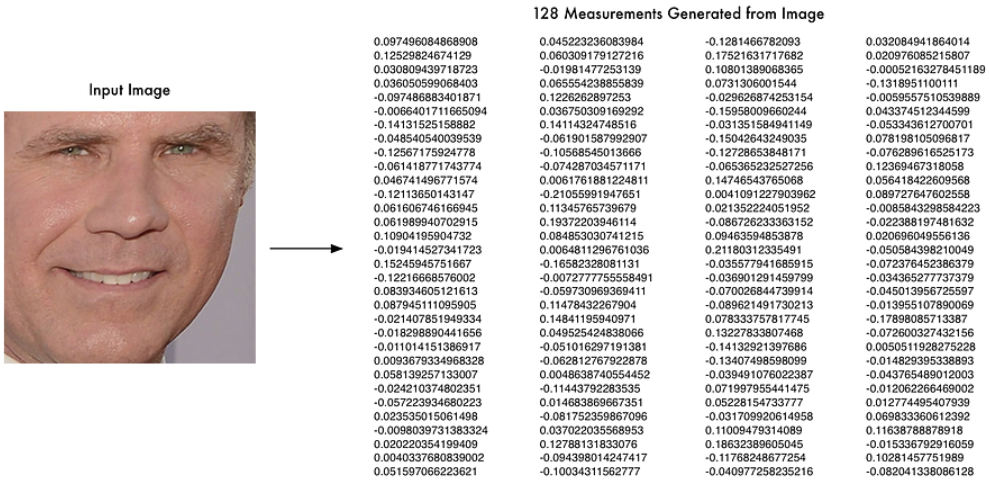


Figure 5.9 128 measurements generated from image

The last step, finding the person in the database with the closest measurements to our test image is actually the easiest step in the entire face recognition process. To accomplish this, we can use a basic machine learning classification algorithm. There is no need for any advanced deep learning techniques, as a simple linear SVM classifier or other classification algorithms can be used. The face-recognition library already do this for us.

5.6 Testing

Unit Testing: This type of testing involves testing individual units or components of the software application in isolation, to ensure that each unit works as intended. In the case of the human detection system, unit testing could involve testing individual functions such as image preprocessing, feature extraction, and classification to ensure that each component works correctly.

Integration Testing: Integration testing is performed to ensure that the individual units of the software application work together as expected, without any issues or errors. For the human detection system, integration testing could involve testing the integration of image preprocessing, feature extraction, and classification to ensure that the overall system works correctly.

Functional Testing: This type of testing verifies that the system functions as expected based on its requirements. For our human detection system, functional testing could involve testing whether the system can accurately detect and identify humans in the surveillance video, as well as whether it can alert security personnel when a human is detected.

System Testing: System testing is conducted to verify that the software application as a whole meets the specified requirements and performs as expected. In the case of the human detection system, system testing could involve testing the entire system end-to-end, from the input video to the output of detected humans, to ensure that the system meets the requirements and performs as expected.

Acceptance Testing: Acceptance testing is performed to determine whether the software application meets the user's requirements and is ready for deployment. In the case of the human detection system, acceptance testing could involve having users test the system and provide feedback on whether it meets their needs and is ready for deployment.

Regression Testing: This type of testing is performed to verify that changes made to the software application, such as bug fixes or feature enhancements, do not introduce new defects or break existing functionality. In the case of the human detection system, regression testing could involve retesting the system after making changes to ensure that it still works correctly.

Performance Testing: Performance testing is conducted to assess the performance and scalability of the software application under various loads and conditions. For the human detection system, performance testing could involve testing the system's performance and scalability under different video resolutions, frame rates, and number of people in the video.

Usability Testing: Usability testing is conducted to assess the ease of use and user-friendliness of the software application. For the human detection system, usability testing could involve testing the system's user interface and interaction design to ensure that it is easy to use and understand.

Security Testing: Security testing is performed to identify and mitigate potential security vulnerabilities and ensure the safety and confidentiality of sensitive data. In the case of the human detection system, security testing could involve testing the system's ability to prevent unauthorized access and protect the privacy of individuals in the video.

6. CONCLUSION

In conclusion, the Human Detection in Surveillance Video system offers a valuable solution to improve the efficiency and accuracy of surveillance investigations. By providing an easy-to-use interface that incorporates facial recognition technology and machine learning algorithms, the system allows for quick identification and tracking of wanted individuals. This can greatly reduce the human effort required in investigations, and help law enforcement agencies and other organizations more effectively protect public safety.

Additionally, the system offers a range of features and customization options, such as the ability to upload photos, videos, and even live webcam feeds for analysis. It also allows administrators to add and manage details of individuals and mark them as wanted, making it an effective tool for both criminal investigations and search and rescue missions. Moreover, the system is designed with privacy in mind, as it does not track individuals unless they are marked as wanted. This helps to address concerns about potential misuse or violation of privacy.

Overall, the Human Detection in Surveillance Video system represents a significant advancement in surveillance technology, offering a more efficient and effective way to detect and identify individuals. With further improvements and developments, it has the potential to become an valuable tool for law enforcement and security agencies worldwide.

FUTURE WORK

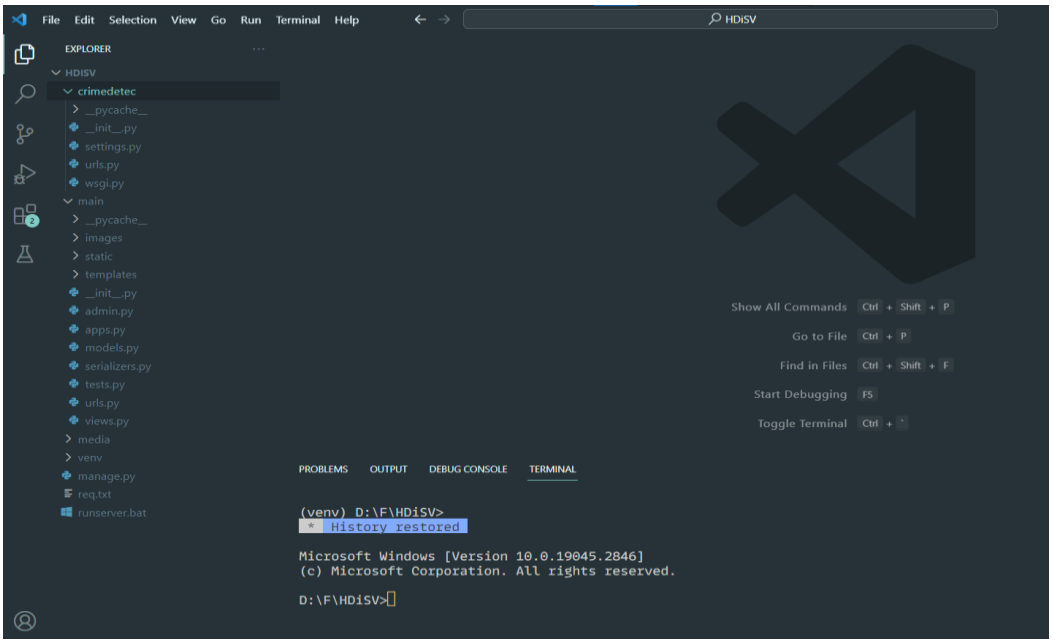
There are several avenues for future work and improvements for the Human Detection in Surveillance Video project. Nothing is perfect and complete and there is always a scope of improvement in each and every product. Everything needs to be updated or upgraded on a timely basis to cope up with the current technology. There are many future scopes for this project. There were many features we had hoped to integrate into our system, but we were compelled to cut them due to time constraints. However, given more time to work on the product, there are a few changes that we would make.

- Integration with more advanced machine learning algorithms to improve the accuracy and speed of detection. This could include incorporating techniques such as object detection and tracking, which would allow for the system to follow a person's movements through multiple frames and provide more detailed information about their actions.
- Integration with other databases, such as criminal records or missing persons databases, to provide more comprehensive information about the individuals detected by the system.
- Currently, the system is designed to work with a single camera. However, you can explore ways to extend the system to work with multiple cameras, which will increase the coverage and efficiency of the system.
- Development of a mobile application that would allow law enforcement officers to access the system's functionality on-the-go, enabling them to quickly and easily identify potential suspects while in the field.

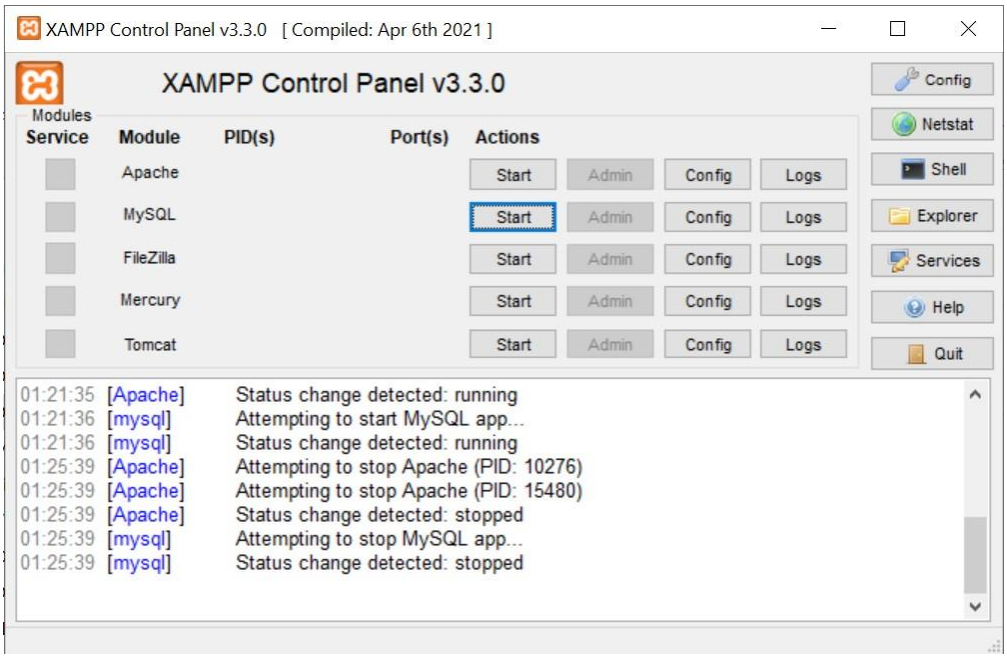
User Manual

1: Open the VS-Code or any other editor for opening the project.

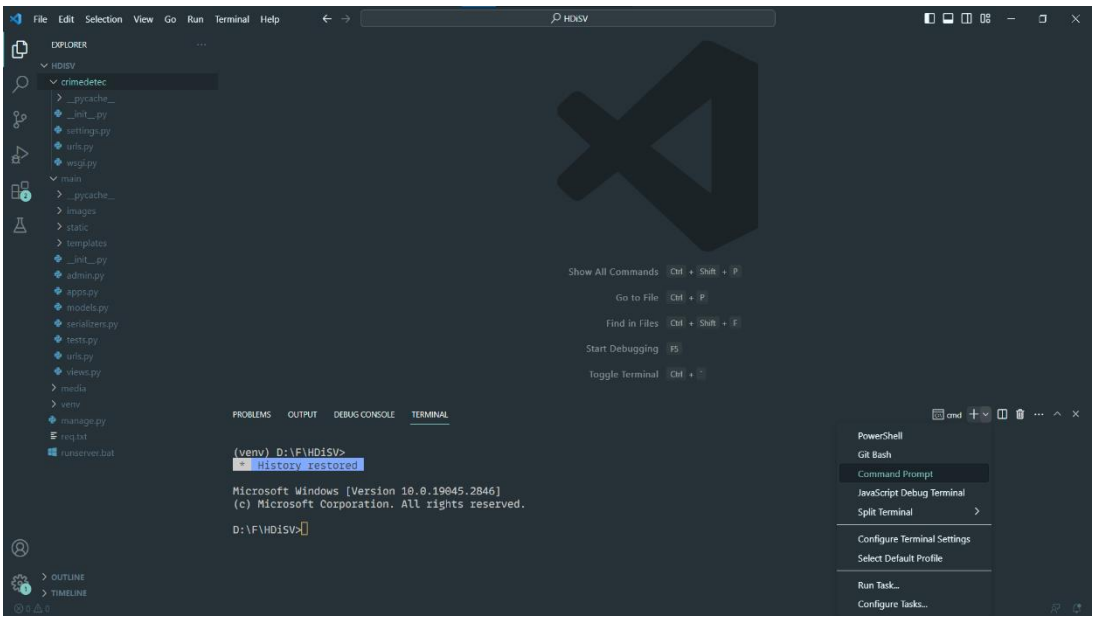
2. Open the folder in which web application files are stored.



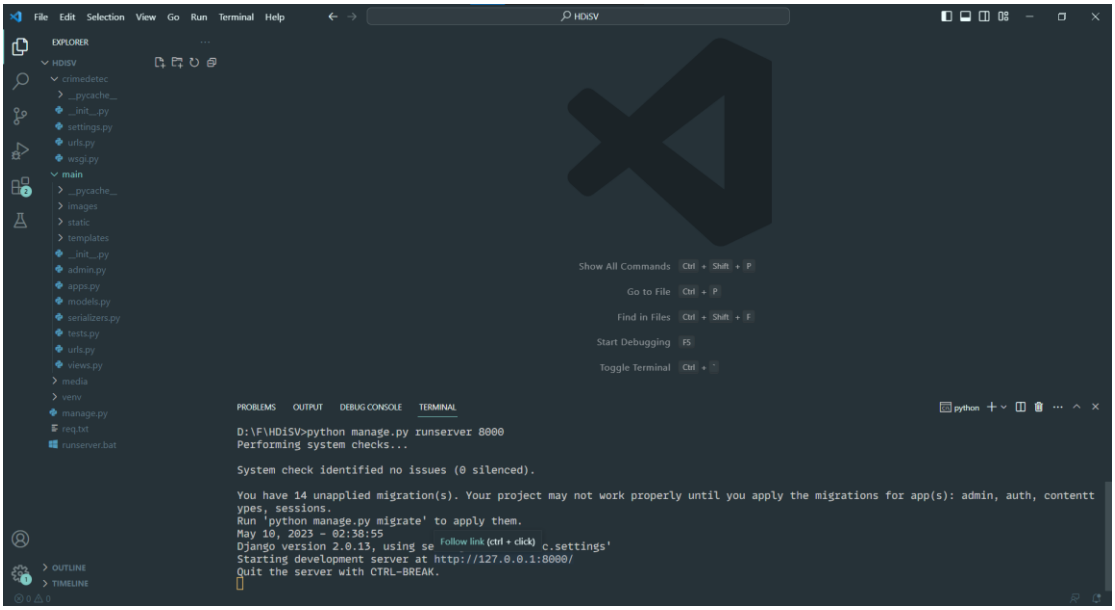
3. Open the XAMPP Control panel and start Apache and MySQL module.

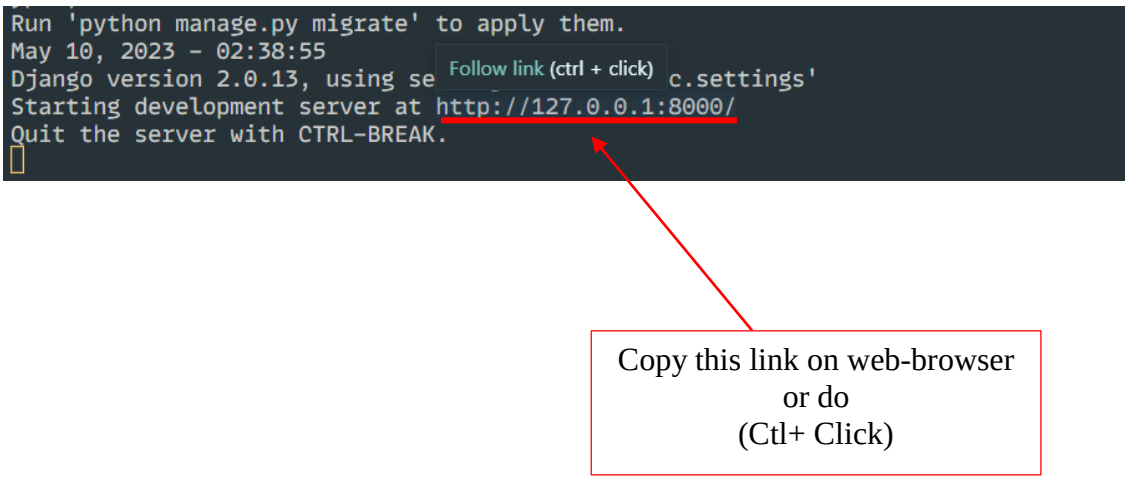


4. Click on the New Terminal and select Command Prompt



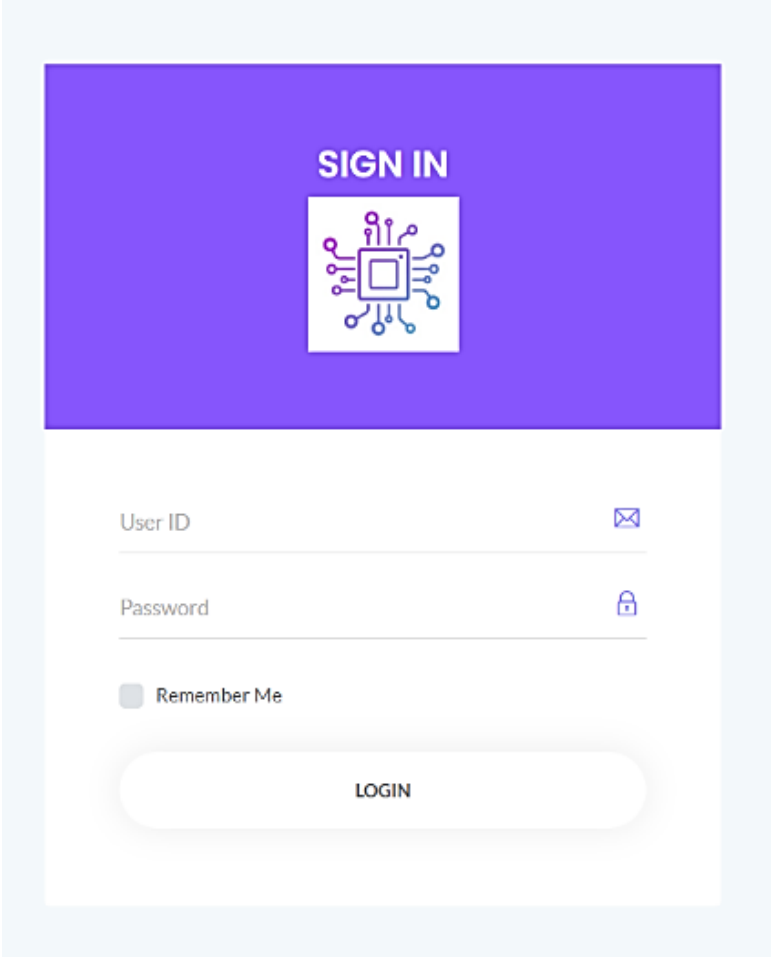
5. Type **python manage.py runserver 8000** on the terminal .You will get following output with link to open web application on local host.



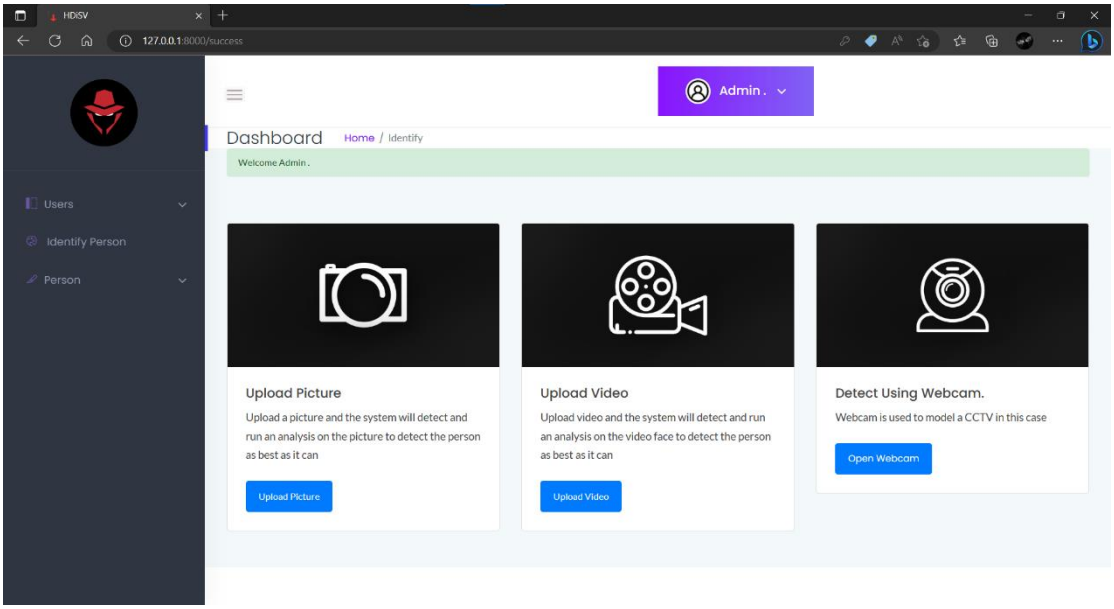


Copy this link on web-browser
or do
(Ctl+ Click)

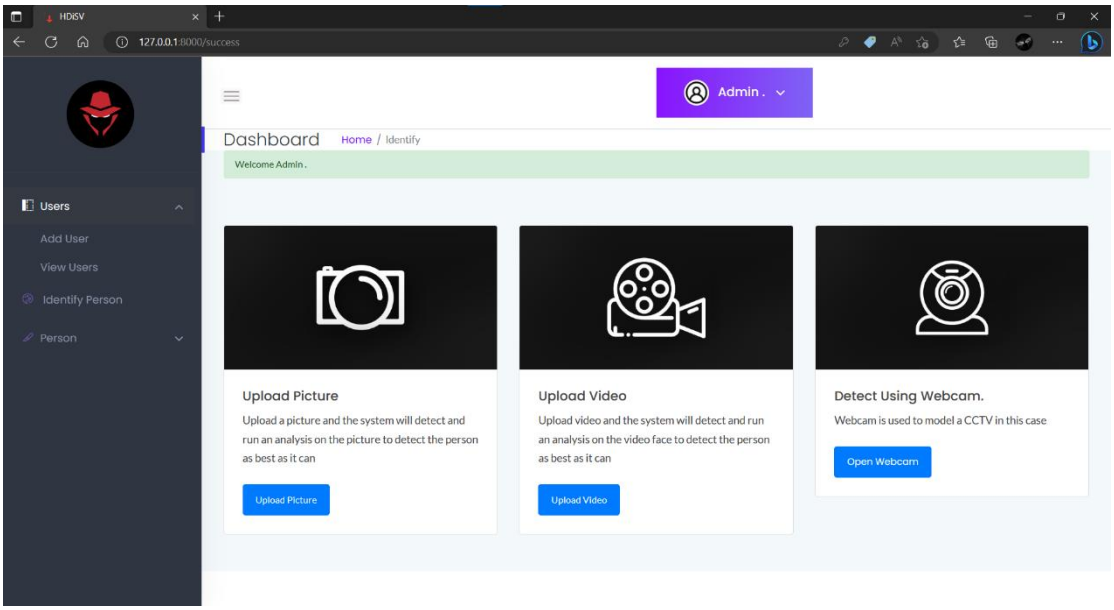
6. After clicking and opening link in web browser, login window will appear. Input the credentials to get access of the system



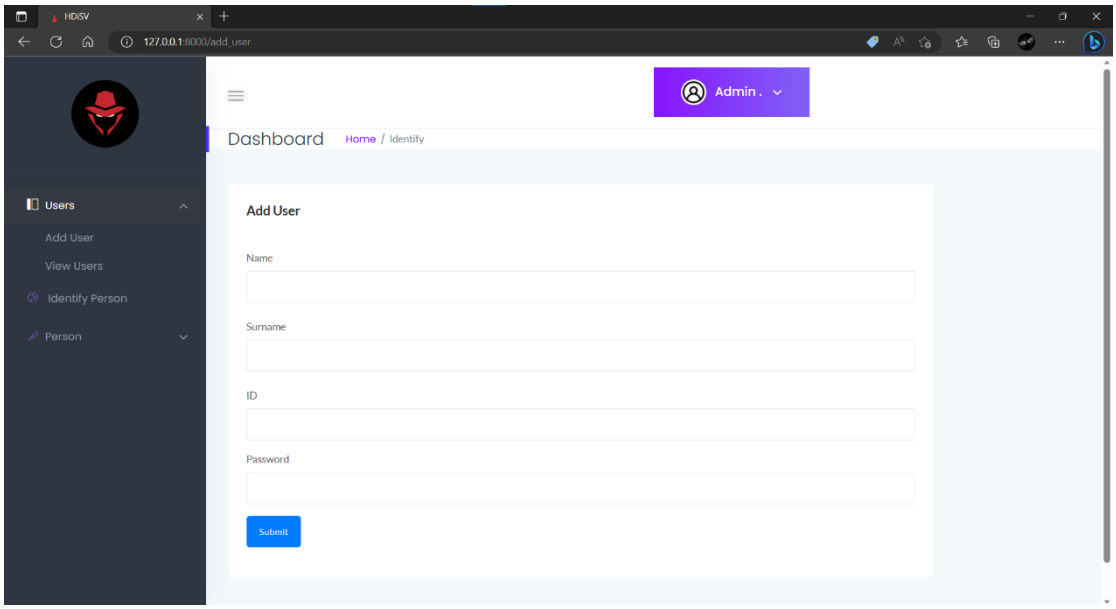
7. After that you will gain access of the system.



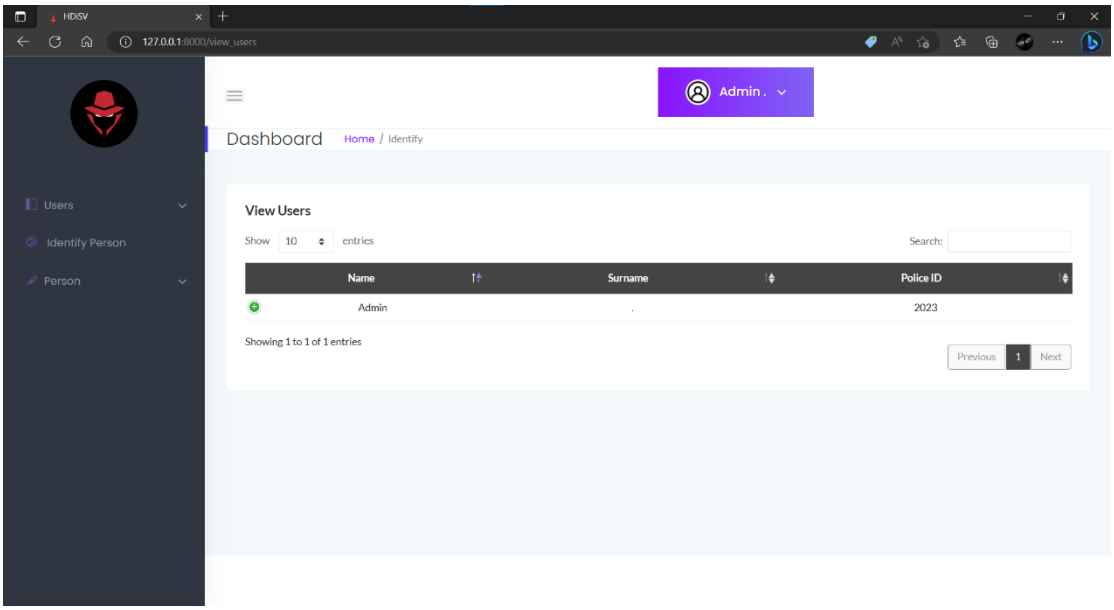
8. After cliking on Users two option will appear, Add user and View Users



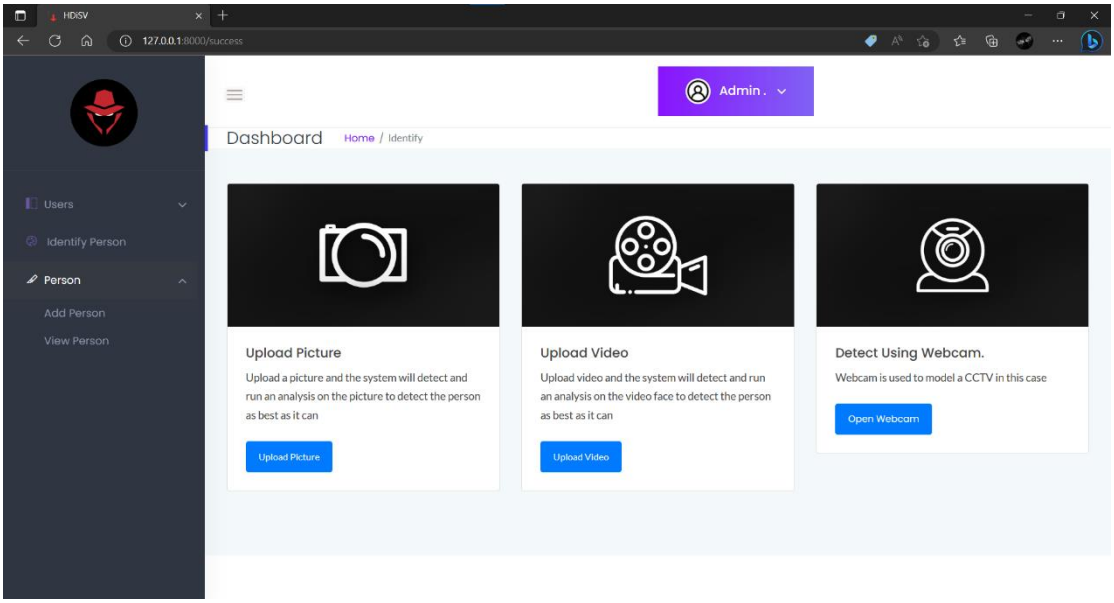
9. New user can be added using following option.



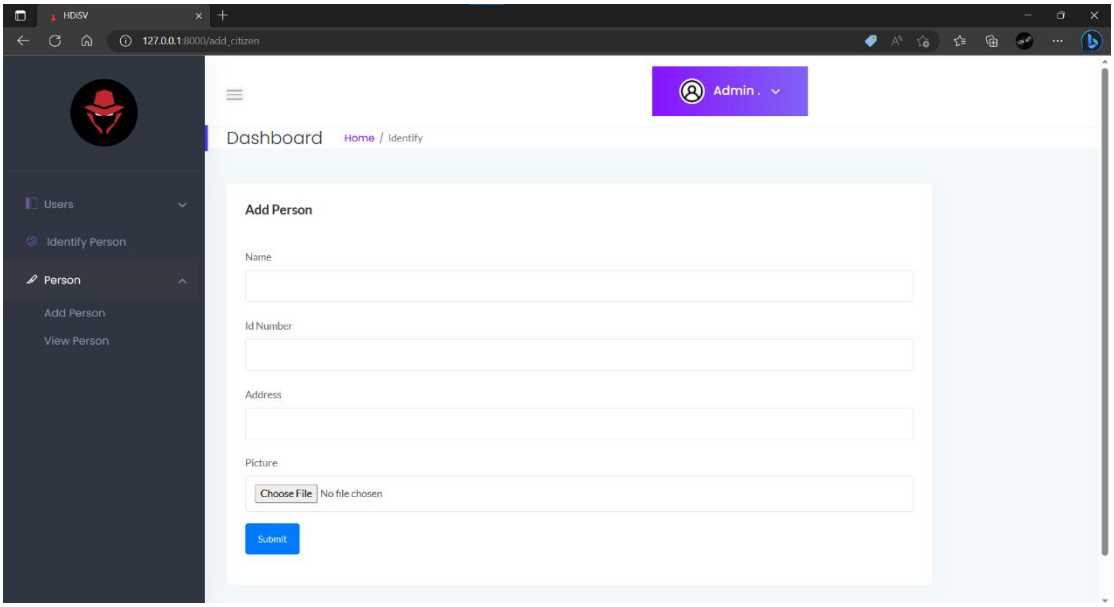
10. System provides functionality to view users.



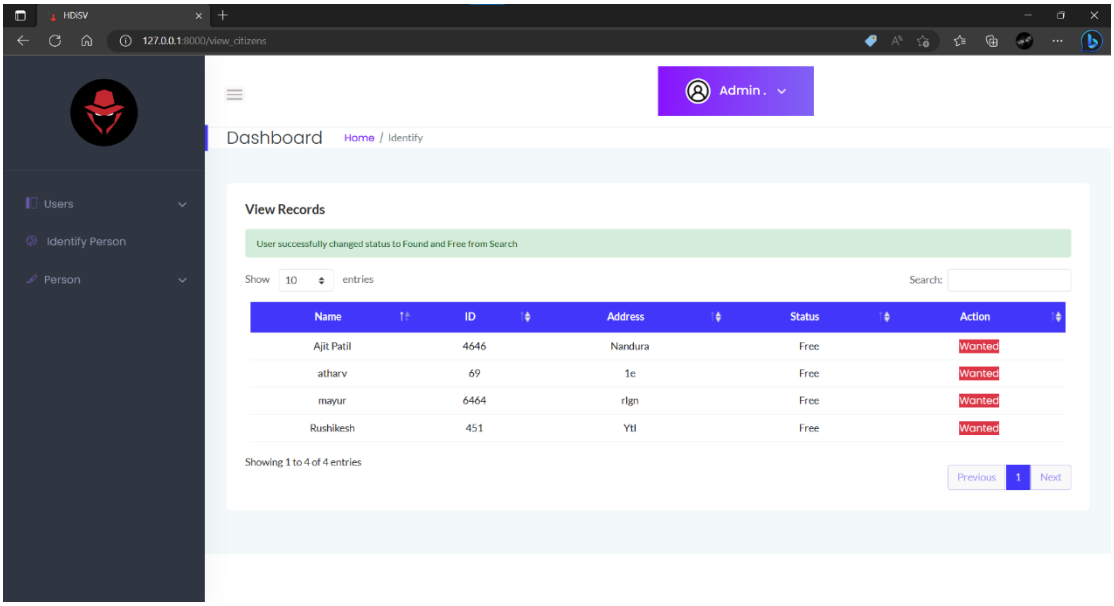
11. After cliking on Person two option will appear, Add Person and View Person.



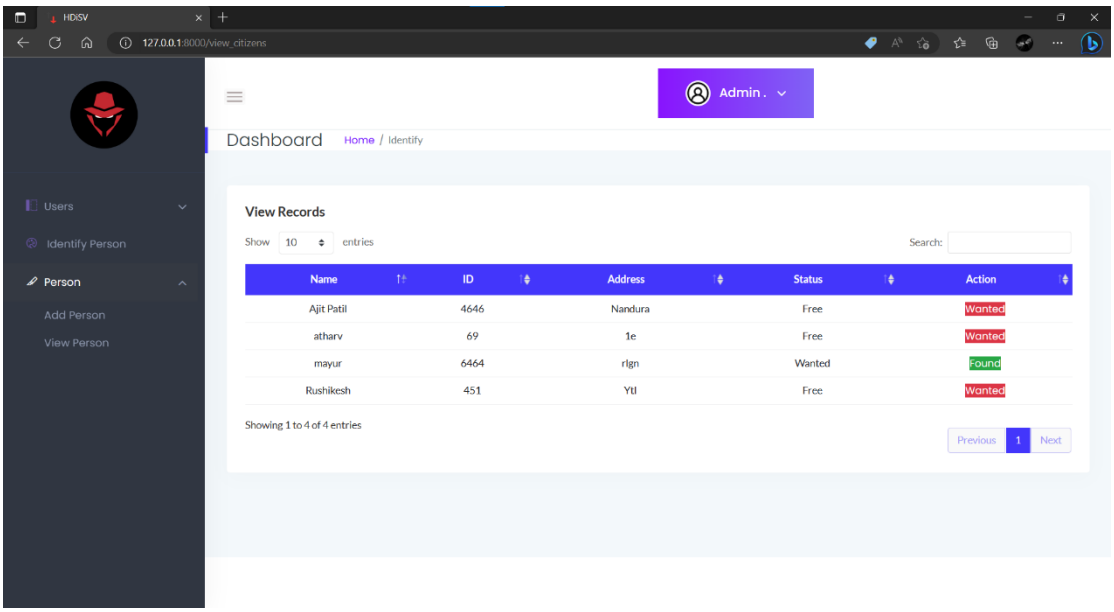
12. User can add person details like Name, Surname, Id , Address along with the photo of that person.



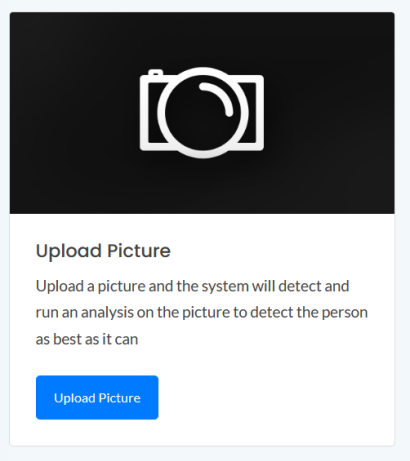
13. User can View Records. This functionality also allow user to mark particular individual wanted or free. It helps in detecting that person during investigation. As it can be seen everyones status is currently free but there is option to mark them as wanted.

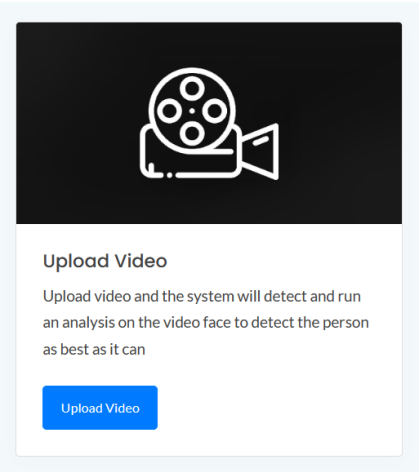


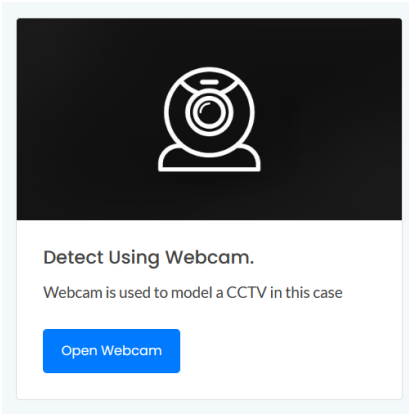
14. After marking someone wanted status changes. As shown below statues of id 6464 changed from free to wanted.



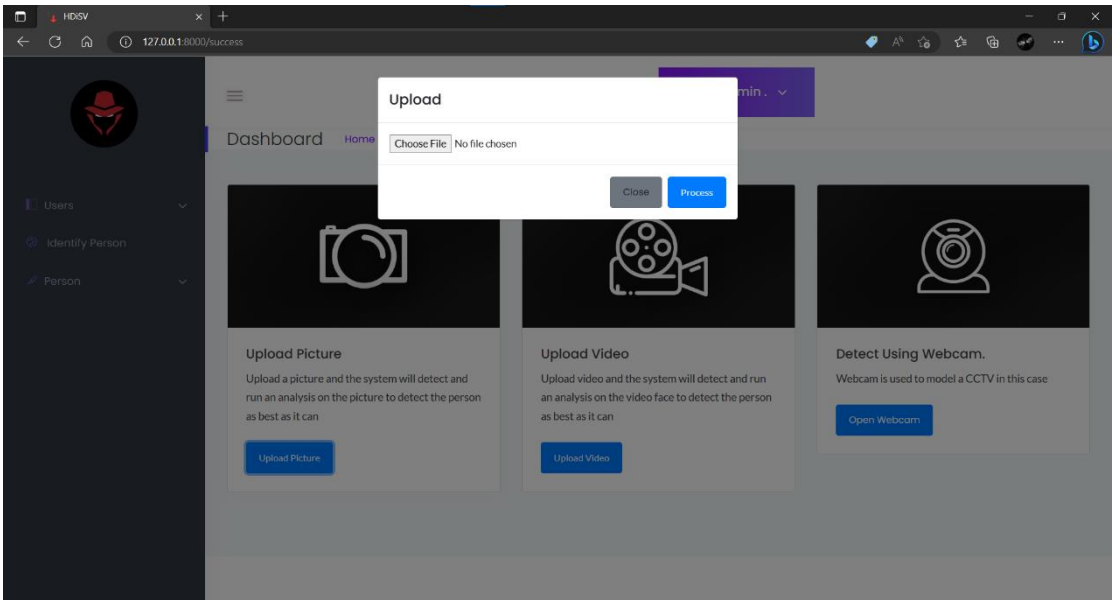
15. On the Dashboard there are three options Upload Picture, Upload Video, Detect Using Webcam. Each option does the same as written in their description.



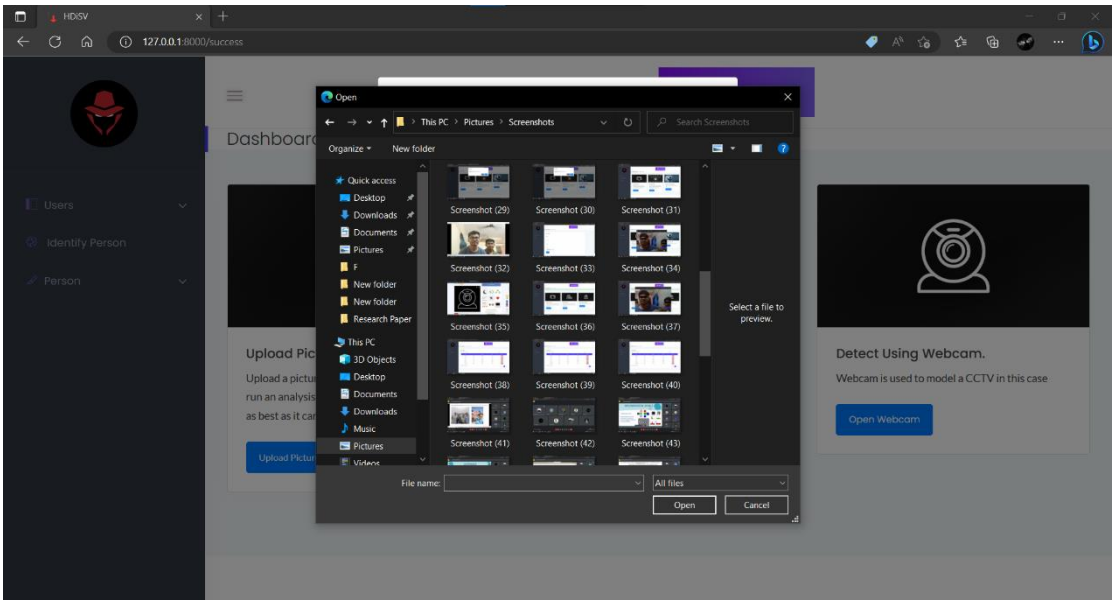




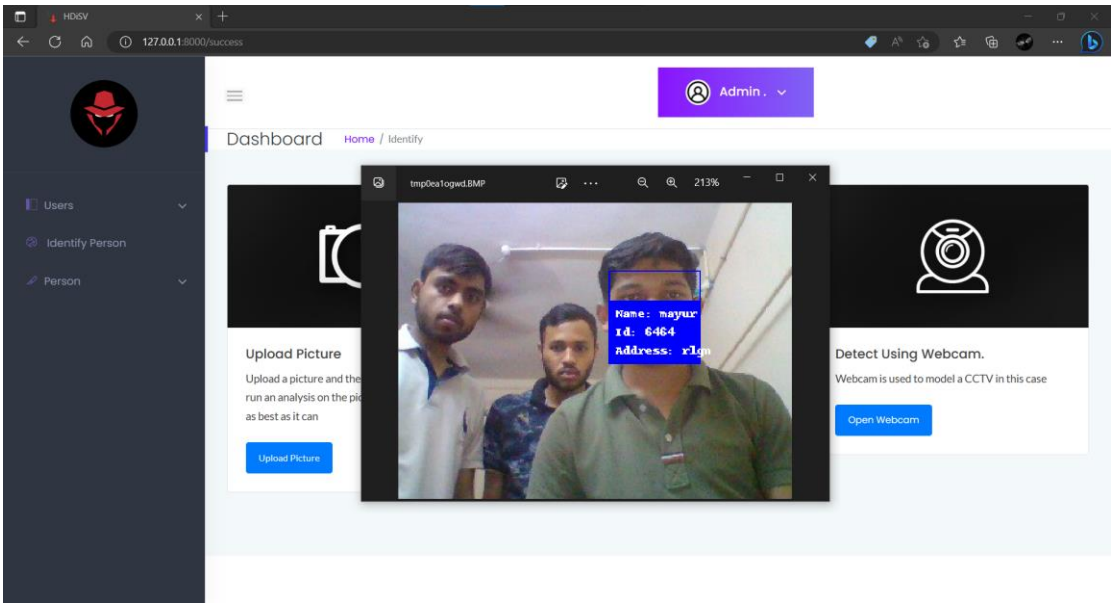
16. After clicking on the Upload Picture option following window will appear. It allows to upload image of crowd from which, a person marked wanted will be detected along with its details.



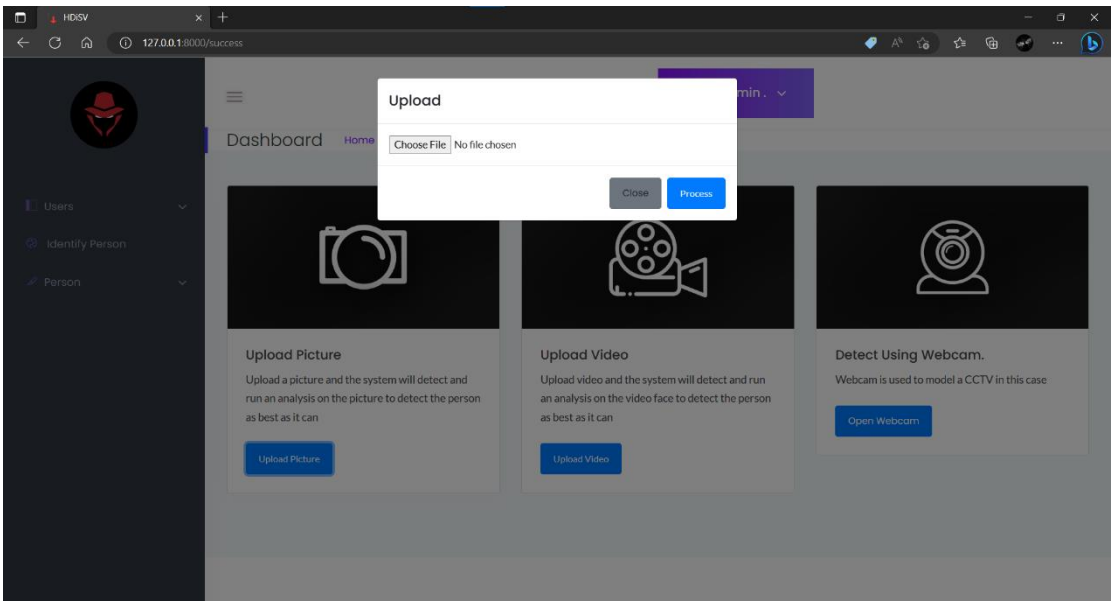
After clicking choose file below window will appear.



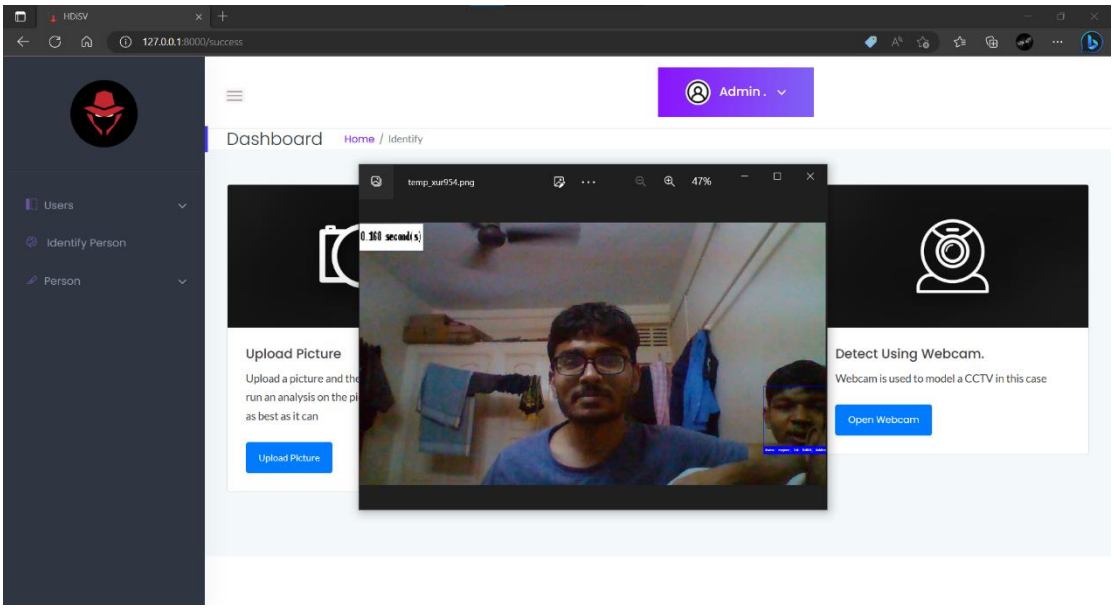
17. Output of Upload Picture option will be an image with detected persons information as shown below.



18. This window will appear after clicking on Upload Video option. It allows user to upload the recorded video from which wanted person is to be detected.



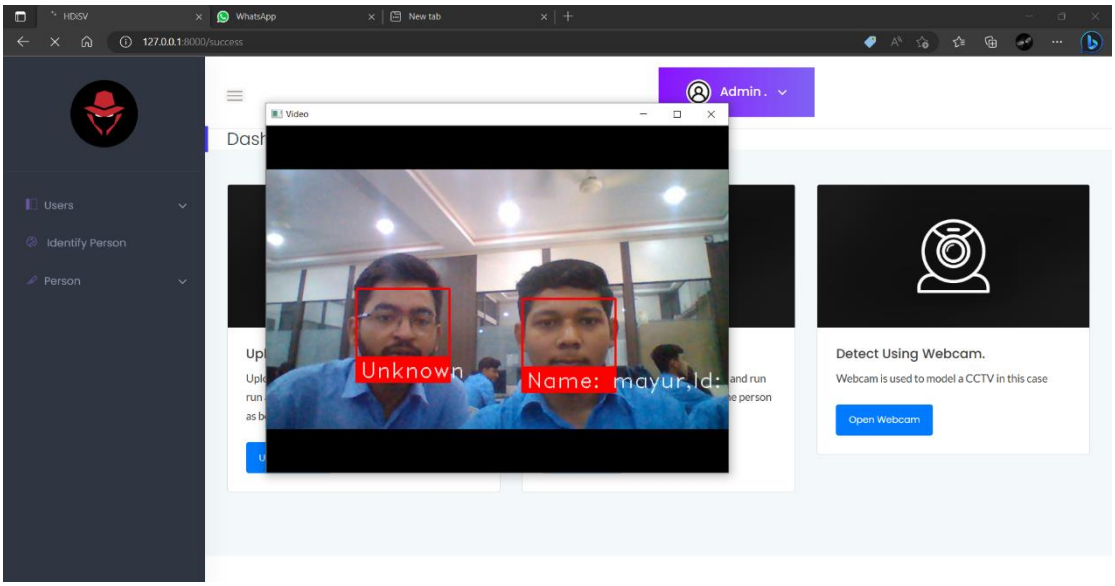
19. Output of Upload Video option will be a video frame with detected persons information and the timestamp of his first appearance as shown below. It helps in reducing the time required during investigations.



Timesta

Bounding box is created around the face of detected person along with

18. This window will appear after clicking on Detect Using Webcam option. Here webcam is used to model CCTV. It provides realtime detection of people and shows their details. If that person is not present in database it will mark them as unknown as shown in the following image.



REFERENCES

[1] Kai Jin, Xuemei Xie , Fangyu Wang, Xiao Han, Guangming Shi, "Human Identification Recognition In Surveillance Videos" 2019, IEEE International Conference on Multimedia & Expo Workshops (ICMEW)

[2] Vandit Gajjar, Ayesha Gurnani, Yash Khandhediya, "Human Detection and Tracking for Video Surveillance: A Cognitive Science Approach" 2017, IEEE International Conference on Computer Vision Workshops (ICCVW)

[3] Mr.R.Prashanth Kumar, Abdul Majeed, Farhan Pasha, A Sujith, "Real-Time Criminal Identification System Based On Face Recognition", NCRENB National Paper Presentation-2023, VIVA-Tech International Journal for Research and Innovation Volume 1, Issue 4 (2023), ISSN(Online): 2581-7280S

[4] Martino C. Khuangga & Dwi H. Widyantoro, "Human Identification Using Human Body Features Extraction" International Conference on Advanced Computer Science and Information Systems (ICACSIS), October 2018

[5] R. Rathi, M. Choudhary & B. Chandra, "An Application of Face Recognition System using Image Processing and Neural Networks", International Journal Computer Technology Application, 3:1, (2012), pp. 45-49.

[6] Mohammed Kabiru Halidu, Pooneh Bagheri-Zadeh, Akbar Sheikh-Akbari, Reinhold Behringer, "PCA in the context of Face Recognition with the Image Enlargement Techniques", 2019 8th Mediterranean Conference on Embedded Computing (MECO), ISSN: 2377-5475

[7] A. K. Jain, Arun A. Ross, Nandakumar, K. "An introduction to biometric recognition", Springer, ISBN978-0-387-77326-1DOI10.1007/978-0-387-77326-1

[8] Edy Winarno, Agus Harjoko, Aniati Murni Arymurthy, Edi Winarko, December, "Face Recognition Based on Symmetrical Half-Join Method using Stereo Vision Camera" 2016, International Journal of Electrical and Computer Engineering 6(6):2818, DOI:10.11591/ijece.v6i6.pp2818-2827

[9] Christian Szegedy, Vincent Vanhoucke, Sergey Ioffe, Jon Shlens, "Rethinking the Inception Architecture for Computer Vision", 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA

[10] G.H. Givens, J.R. Beveridge, P.J. Phillips, B. Draper, Y.M. Lui, D. Bolme, "Introduction to face recognition and evaluation of algorithm performance" Computational Statistics & Data Analysis,Volume 6701, November 2013, pp 236–247

[11] M. Parisa Beham, S. Mohamed Mansoor Roomi, "A Review Of Face Recognition Methods", April 2013, International Journal of Pattern Recognition and Artificial Intelligence 27(4), DOI: 10.1142 /S0218001413560053

[12] I. Haritaoglu, D. Harwood and L. S. Davis, "Real-time surveillance of people and their activities", IEEE Trans. Pattern Anal. Mach. Intell. 22(2000) 809–830.

[13] Mr. Pankaj S. Parsania1 , Dr. Paresh V. Virparia, "A Comparative Analysis of Image Interpolation Algorithms", January 2016, IJARCCE 5(1):29-34, DOI: 10.17148/IJARCCE.2016.5107

[14] Zhifei Wang, Zhipeng Li, Lan Cheng, Gaowei Yan, "An improved ORB feature extraction and matching algorithm based on affine transformation", Conference: 2020 Chinese Automation Congress (CAC), DOI:10.1109/CAC51589.2020.9327165

[15] Ondrej Miksik CMP and Krystian Mikolajczyk, "Evaluation of Local Detectors and Descriptors for Fast Feature Matching", Pattern Recognition (ICPR), 2012 21st International Conference

[16] Ethan Rublee, Vincent Rabaud , Kurt Konolige and Gary Bradski, "ORB: an efficient alternative to SIFT or SURF" 2011 International Conference on Computer Vision, Date of Conference: 06-13 November 2011

[17] Michael Calonder, Vincent Lepetit, Christoph Strecha, and Pascal Fua "BRIEF: Binary Robust Independent Elementary Features", European Conference on Computer Vision ECCV 2010: Computer Vision – ECCV 2010,pp 778–792

[18] H. Bay, A. Ess, T. Tuytelaars and L. V. Gool "SURF: Speeded Up Robust Features" , European Conference on Computer Vision ECCV 2006: Computer Vision- ECCV 2006, pp 404–417

[19] Edward Rosten, Thomas Drummond, "Machine Learning for High-Speed Corner Detection", European Conference on Computer Vision ECCV 2006: Computer Vision – ECCV 2006 pp 430–443

[20] Joseph Redmon, Santosh Divvala, Ross Girshick, Ali Farhadi, "You Only Look Once: Unified, Real-Time Object Detection" , 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), DOI: 10.1109/CVPR.2016.91

[21] Joseph Redmon, Ali Farhadi, "YOLO9000: Better, Faster, Stronger", Computer Vision and Pattern Recognition, Cite as: arXiv:1612.08242

[22] Chandan G, Ayush Jain, Harsh Jain, Mohana, "Real Time Object Detection and Tracking Using Deep Learning and OpenCV", 2018 International Conference on Inventive Research in Computing Applications (ICIRCA), DOI: 10.1109/ICIRCA .2018.8597266

[23] Panca Mudjirahardjo and Hadi Suyono, "Real Time Boundary Density based Moving Object Extraction", 2018 3rd Technology Innovation Management and Engineering Science International Conference (TIMES-iCON), DOI: 10.1109/TIMES-iCON.2018.8621754

[24] Ross Girshick, Jeff Donahue, Trevor Darrell, Jitendra Malik, "Rich feature hierarchies for accurate object detection and semantic segmentation", 2014 IEEE Conference on Computer Vision and Pattern Recognition, DOI: 10.1109/CVPR. 2014.81

[25] Christian Szegedy, Alexander Toshev, Dumitru Erhan, "Deep Neural Networks for Object Detection", January 2013 Conference: Advances in Neural Information Processing Systems

[26] Thomas L. Dean, Mark A. Ruzon, Mark E. Segal, Jonathon Shlens, Sudheendra Vijayanarasimhan, J. Yagnik, "Fast, Accurate Detection of 100,000 Object Classes on a Single Machine", 2013 IEEE Conference on Computer Vision and Pattern Recognition, DOI:10.1109/CVPR.2013.237 Corpus ID: 2568065

[27] Ebubekir BUBER, Banu DIRI, "Performance Analysis and CPU vs GPU Comparison for Deep Learning", 2018, 6th International Conference on Control Engineering & Information Technology (CEIT), DOI: 10.1109/CEIT.2018.8751930

[28] Shuo Liu, Xiang Peng and Zheng Liu, "Image Quality Assessment through Contour Detection", 2019 IEEE 28th International Symposium on Industrial Electronics (ISIE), doi:10.1109/ISIE.2019.8781416

[29] T.-J. Liu, W. Lin, and C.-C. Kuo, " Image Quality Assessment Using Multi-Method Fusion" 2012 IEEE Transactions on Image Processing, PP:1793-807. doi: 10.1109/TIP.2012.2236343.

[30] Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli,"Image quality assessment: From error visibility to structural similarity" by, 2004 IEEE Transactions on Image Processing, Volume: 13, Issue: 4, April 2004, Page(s): 600 - 612

[31] Z. Wang and A. C. Bovik, "A universal image quality index", 2004 IEEE Signal Processing Letters (Volume: 9, Issue: 3, March 2002) ,Page(s): 81 - 84

[32] X. Zhang, X. Feng, W. Wang, and W. Xue, "Edge strength similarity for image quality assessment" April 2013, Signal Processing Letters, IEEE 20(4):319-322

[33] Vahid Kazemi and Josephine Sullivan, "One Millisecond Face Alignment with an Ensemble of Regression Trees" Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2014, pp. 1867-1874

DISSEMINATION OF WORK

Research Paper

Title - Human Detection in Surveillance Video

Author - Ajay Dandge, Ajit Patil, Mayur Nehare, Hrushikesh Shukla

Publisher - International Journal of Scientific Research in Engineering and Management (IJSREM)

ISSN No - 2582-3930

DOI - 10.55041/IJSREM20351

Issue – May 2023

Certificates of Participation


"Techno Social Excellence"
Marathwada Mitra Mandal's
Institute of Technology (MMIT), Pune
Accredited with "A" Grade by NAAC
Approved by AICTE | Recognized by DTE | Affiliated to SPPU Pune.
"येथे बहुतांचे हित"
NAAC
Techno-Sci 2K23
Certificate of Participation
Ajit Shivajirao Patil
Has Participated in the Technical Event 'Project Competition' Techno-sci 2K23 held at Marathwada Mitra Mandal's Institute of Technology. Lohgaon, Pune During 25th to 26th April 2023
Prof. S. S. Chaudhari
Event Coordinator
MMIT
INSTITUTION'S
INNOVATION
COUNCIL
Tech Yuva
Dr. R. V. Bhortake
Principal
UADFYA-CE000190
Made for free with Certify'em


"Techno Social Excellence"
Marathwada Mitra Mandal's
Institute of Technology (MMIT), Pune
Accredited with "A" Grade by NAAC
Approved by AICTE | Recognized by DTE | Affiliated to SPPU Pune.
"येथे बहुतांचे हित"
NAAC
Techno-Sci 2K23
Certificate of Participation
Hrushikesh Shukla
Has Participated in the Technical Event 'Project Competition' Techno-sci 2K23 held at Marathwada Mitra Mandal's Institute of Technology. Lohgaon, Pune During 25th to 26th April 2023
Prof. S. S. Chaudhari
Event Coordinator
MMIT
INSTITUTION'S
INNOVATION
COUNCIL
Tech Yuva
Dr. R. V. Bhortake
Principal
UADFYA-CE000128
Made for free with Certify'em


"Techno Social Excellence"
Marathwada Mitra Mandal's
Institute of Technology (MMIT), Pune
Accredited with "A" Grade by NAAC
Approved by AICTE | Recognized by DTE | Affiliated to SPPU Pune.
"येथे बहुतांचे हित"
NAAC
Techno-Sci 2K23
Certificate of Participation
Mayur Nehare
Has Participated in the Technical Event 'Project Competition' Techno-sci 2K23 held at Marathwada Mitra Mandal's Institute of Technology. Lohgaon, Pune During 25th to 26th April 2023
Prof. S. S. Chaudhari
Event Coordinator
MMIT
INSTITUTION'S
INNOVATION
COUNCIL
Tech Yuva
Dr. R. V. Bhortake
Principal
UADFYA-CE000123
Made for free with Certify'em

SOURCE CODE LISTINGS

1. templates –
    - add_user.html
    - add_person.html
    - reports.html
    - upload_pic.html
    - view_persons.html
    - view_users.html
    - welcome.html
2. admin.py
3. apps.py
4. models.py
5. urls.py
6. views.py
7. manage.py

INFORMATION OF MEMBERS

Name: Ajay Dandge

Email: ajaydandge12@gmail.com

Mobile: 9970404649

Address: Shivshankar nagar, Buldhana-443001

Name: Ajit Patil

Email: patilajit2206@gmail.com

Mobile: 9307325176

Address: Plot No. 40B, Khadse Plots, Nandura-443404

Name: Hrushikesh Shukla

Email: hrishabshukla07@gmail.com

Mobile: 9309587724

Address: Yavatmal-445001

Name: Mayur Nehare

Email: mayur.nehare01@gmail.com

Mobile: 9021526912

Address: Ralegaon, Yavatmal-445402