REAL TIME FACE DETECTION AND PERSON TRACKING FOR SURVEILLANCE

By

Zainab Iftikhar	**01-132162-048**
Muhammad Imran	**01-132162-011**
Muhammad Haseeb Akram	**01-132162-010**

Supervised by

Ma'am Tooba Khan

Department of Computer Engineering

Bahria University Islamabad.

2020

REAL TIME FACE DETECTION AND PERSON TRACKING FOR SURVEILLANCE

By

Zainab Iftikhar	**01-132162-048**
Muhammad Imran	**01-132162-011**
Muhammad Haseeb Akram	**01-132162-010**

Supervised by

Ma’am Tooba Khan

Submitted to the department of computer engineering in the partial fulfillment of the requirements for the degree of Bachelors in Computer Engineering.

Department of Computer Engineering

Bahria University Islamabad.

2020

CERTIFICATE

Bahria University Islamabad
Department of Computer Engineering

Dated: 15-07-2020

CERTIFICATE

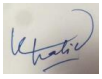
We accept the work contained in the report titled "Real-time Face Detection and Person Tracking for Surveillance" as a confirmation to the required standard for the partial fulfillment the degree of Bachelor of Computer Engineering.

Mr. Waleed Manzoor
Project Coordinator
Date: **15-07-2020**

Engr. Tooba Khan
Supervisor
Date: **15-07-2020**

Engr. Waleed Manzoor
Internal Examiner
Date: **15-07-2020**

Muhammad Imran
External Examiner
Date: **15-07-2020**



Head of Department (CE)
Date: **15-07-2020**

UNDERTAKING

We Zainab Iftikhar, Muhammad Imran, and Muhammad Haseeb Akram certify that a project titled *"Real-Time Face Detection and Person Tracking for Surveillance"* is our own work. The work given to us has not been presented elsewhere for assessment. Material that has been used form the other sources has been properly acknowledged.

Zainab Iftikhar (47170)

Muhammad Imran (47139)

M.Haseeb Akram (47138)

DEDICATION

We would like to dedicate this project to Allah Almighty, to our beloved parents and supervisor Miss Tooba Khan, who help and encourage us to choose this project and help us throughout the whole project. We are also thankful to our families who support us and stay with us through thick and thin problems.

Last but not the least, we dedicate this thesis to our friends who support us and delighted us when we exhausted and tired of the work.

ACKNOWLEDGMENTS

We Ms. Zainab Iftikhar, Mr. Muhammad Imran, Mr. Muhammad Haseeb Akram are most thankful to the Allah Almighty. With blessings of Allah Almighty, we can complete our Final Year Project. We would like to express sincere gratitude to our supervisor “Miss Tooba Khan” for her motivation, support, guidance, and immense knowledge. Her guidance helped us in the thesis and throughout the whole project. We couldn’t have imagined such a great mentor who encourages us to avail of better opportunities.

We would also like to thank all our fellows who helped us in doing a project. Our sincere appreciation to the Head of the supervisor and to all committee members who always encourage us and make us prepare and perform as efficiently.

Last but not least we are also thankful to our families who support us and stay with us through thick and thin problems. They are the superlative role models. Our sincere gratitude to all.

ABSTRACT

Nowadays there are security systems that are extensively used for security, surveillance, observing and monitoring person in different places. Completely automated systems are competent of analyzing activities and information in an image or a video by detecting face, recognize and track the person. Person tracking track the same person across multiple cameras or network of cameras. These systems are required for security purposes in organizations. Specific application includes access control to secure environments, identification of individuals at a place and intruder detection. To able to provide a secure environment, a high security system will be required. Our project "Real time face detection and person tracking for surveillance" will provide security system which contains face detection and person tracking which allows an organization to keep a track on all its employees.

For Real-time Person tracking, the first stage is to detect the person's face using a single camera or multiple cameras. For face detection, we must preprocess the images. Image processing done by converting RGB images into grayscale, and for enhancing image we use contrast stretching. Pre-process images to remove noise and blurriness in images and system detect the person face in an image. After face detection, we trained our model to extract facial features and stored in the database.Person tracking done by face recognition, we apply Haar cascade classifier for face Recognition. We used location and time stamps for the person tracking which will show us the time when the person has passed from the camera and also tells us the current location of a person and camera number. Our Tracking system is very usefull for security purposes in organization where movement of persons shold be tracked to keep an eye on movement of every person in the organization for security purpose. Our system will be best for highly secured orgainzation to keep a track of all its employees that where a person is going on which time and which place.

Keywords:

Artificial Intelligence (AI), Machine Learning (ML), Computer Vision (CV), Face Detection, Person Recognition, Person Tracking, surveillance, Real-time, Haar Cascade Classifier.

Table of Contents

CERTIFICATE ..... **iii**
UNDERTAKING ..... **iv**
DEDICATION ..... **v**
ACKNOWLEDGMENTS ..... **vi**
ABSTRACT ..... **vii**
LIST OF FIGURES ..... **x**
LIST OF TABLES ..... **xi**
CHAPTER 1 INTRODUCTION **1**
1.1 Artificial Intelligence 2
1.2 Machine Learning 2
1.3 Neural Networks 4
1.4 Convolution Neural Networks: 5
1.5 Deep Learning 8
1.6 Problem Statement: 9
1.7 Objectives 9
1.8 Scope 10
CHAPTER 2 LITERATURE REVIEW **11**
2.1 Face Detection: 11
2.1.1 Feature Based Method *11*
2.1.2 Knowledge Based Method *12*
2.1.3 Template Matching Method *12*
2.1.4 Appearance Based Method *12*
2.2 Face Recognition 13
2.2.1 Holistic Approach Method *14*
2.2.2 Feature Based Matching *14*
2.2.3 Hybrid Face Recognition Method *14*
2.3 Person Tracking 15
2.3.1 Previous work: *15*
2.3.2 Deep Sort: *15*
2.3.3 Our approach: *15*
CHAPTER 3 DESIGN & METHODOLOGY **17**
3.1 Images Preprocessing 18
3.1.1 Grayscale Image *18*
3.1.2 Contrast Stretching *18*
3.2 Import Modules 18
3.3 Face Detection 19
3.3.1 Haar Cascade *19*
3.4 Camera Connectivity 22
3.5 Dataset 22

3.5.1 Packages Installation 23
3.5.2 Labeling Dataset 23
3.6 FACE RECOGNITION 24
3.7 PERSON TRACKING 25
CHAPTER 4 RESULTS & DISCUSSION 26
4.1 IMAGE ACQUISITION 27
4.1.1 Conversion of RGB Image to Grayscale 27
4.1.2 Normalization 27
4.2 FACE DETECTION 28
4.2.1 Eyes Detection 28
4.3 FACE RECOGNITION 29
4.4 PERSON TRACKING USING SINGLE CAMERA 29
4.5 PERSON TRACKING USING MULTIPLE CAMERAS 29
4.5.1 Intruder Detection 30
4.6 ACCURACY 31
4.7 NEGATIVE RESULT 31
4.8 SYSTEM IMPLEMENTATION 31
CHAPTER 5 CONCLUSION 32
CHAPTER 6 FUTURE WORK 33
REFRENCES 34

LIST OF FIGURES

Number *Page*

Figure 1.1 Image Classification Inter-class Variations 6
Figure 1.2 Working of image matrices with the kernel in CNN 7
Figure 1.3 Object Detection and Classification using CNN 7
Figure 1.4 Relationship of subdomains of AI 8
Figure 2.1 Face Detection Techniques 11
Figure 2.2 Face recognition techniques 13
Figure 3.1 Block Diagram 17
Figure 3.2 Face Detection Pipeline 19
Figure 3.3 Architecture of Haar Cascade Classifier 20
Figure 3.4 Face Detection using Haar Cascade Classifier 21
Figure 3.5 Integral Image 21
Figure 3.6 Hercules Duplex HD webcam 22
Figure 3.7 Face Images Dataset 22
Figure 3.8 Packages Installation 23
Figure 3.9 Labelled Face Images 24
Figure 3.10 Face Recognition Pipeline 24
Figure 3.11 Person Tracking Block Diagram 25
Figure 4.1 Image Acquisition 27
Figure 4.2 Conversion from RGB Image to Grayscale 27
Figure 4.3 Normalized Image 277
Figure 4.4 Real-Time Face Detection 28
Figure 4.5 Real-Time Eyes Detection 288
Figure 4.6 Face Recognition 29
Figure 4.7 Person Tracking Using Single Camera 29
Figure 4.8 Tracking Person Time 29
Figure 4.9 Person tracking using Multiple cameras 30
Figure 4.10 Tracking person's time using multiple cameras 300
Figure 4.11 Intruder Detection 300
Figure 4.12 False Output 31
Figure 4.13 System Implementation 31

LIST OF TABLES

Table Number *Page*

Table 2.1 Summary of Earlier Techniques Used 16
Table 4.1 Accuracy of recognition and tracking..... 316

CHAPTER 1

1 INTRODUCTION

Computer vision is a field of computer science which include methods for acquiring, analyzing and processing the images. Computer vision play vital role in surveillance systems, it helps to detect and track the person by face recognition [1]. The demand of cameras for application in surveillance for security purpose is growing rapidly. The real time person tracking and identification for surveillance system allow tracking and recognize person from video or an image and provide an efficient and cost-effective mechanism of securing sensitive areas [2].There is an increasing demand for an application that support and monitor person for both indoor and outdoor environment [3]. Face detection and recognition is today's research problem because it is spread over in many fields. To detect a single or multiple faces P. Viola proposed Viola Jones method [4]. Facial recognition is for the identification of the living person through the face. Now the time has changed, and the things are shifted to the real-time settings. It is developing so fast due to the availability of compatible systems that are required for these works to be implemented [5]. In recent years, attention is increased for this work because it has importance in applications such as surveillance, activity analysis of the specific person, or tracking of the same person with the help of multiple cameras.

Although different methods have been proposed to solve these problems, difficulties remain the same [4]. Sometimes due to the background area and due to the problem of light, it's difficult to recognize the person. So, there is a problem if a variation between the test frame and trained images. For accuracy, a large dataset is required to resist this problem [6]. However, there is the rapid advancement of facial recognition due to the presence of large databases of facial images and there are different methods to calculate the performance of facial recognition algorithms. This way we can find which algorithms are working efficiently or have accuracy [7].

All algorithms consist of two parts:

- Face Detection and Recognition.

- Person Tracking

Such algorithms only give facial images while partially automatic algorithms give the facial images and coordinates of the eyes.

1.1 Artificial Intelligence

Human intelligence demonstrated by the machines is known as Artificial intelligence. The approach of AI includes searching algorithms, machine learning algorithms, problem solving techniques and statistical research. the main features of AI include planning, understanding language, object detection, learning and problem solving. AI is further divided into two parts general AI and narrow AI. Artificial Intelligence involves acting on the data, learn from the new data, and improve it over time [8].

Artificial intelligence basically 'trains' program for a specific task and allows it to explore. To apply Artificial Intelligence, you just need the data and some information about that data. Artificial Intelligence (AI) algorithms are trained on large data datasets so that they can make predictions, identify the patterns, and recommend some actions that much like a human would, just faster and better.

1.2 Machine Learning

It is the sub-field of artificial intelligence in which computer systems learn from examples or given datasets or experience. The software was a program that could play checkers and improved with each game it played. Learning algorithms can accomplish the functions of high complexity. The learning elements generate the systems that can be versatile, flexible and that continue to improve the accuracy of outcomes.

Machine Learning vitality converges in the field of computer science, data science, and data analysis. learning elements used in machine learning for exploring data analysis, for processing, detecting, and predicting future work and learn from patterns to decide. For image related tasks we must first consider the basics of machine learning. To reduce human efforts in a real-world problem and replace that effort by learning algorithms, machine learning was evolved.it has been a helpful tool for problem-solving. With the rapid increase in modern technology and tools, data and computational capacity have also been increased. Thus, machine learning has become more realistic. We will further use its algorithms in the application of computer vision in image classification.

Classification of Machine Learning methods are given below:

In **Supervised assimilation** system is pretrained with an example of a dataset with given labels. Labels show the classes of the given dataset. A learning algorithm is used to train the system then new inputs that are unknown to the system are tested to get the accurate results. Supervised Learning corresponds to human learning under the guidance of the mentor.

Machine learning problems are solved based on two techniques i.e. Classification and regression. Regression problems are determining the relation between targeting the numerical values and variables for example the average price of a house in a certain area can be predicted using regression technique. The classification technique is used to identify different categories and it targets the quantitative variables. The most common example of classification is the prediction of tumors that whether the tumor is malignant or benign.

In **Unsupervised assimilation** In an Unsupervised assimilation algorithm learns from the dataset that is without labels. Algorithm determines the data patterns on its own. This kind of learning is similar to the methods i.e. observation used by human for evaluation of some objects and developments of same class. The technique that is used to solve the Unsupervised Learning problem is clustering. Recommended systems are examples of Unsupervised learning. With the evolution of deep learning now supervised learning problems can be processed using unsupervised learning for revealing the effective illustration of data.

Reinforcement Learning: It lies in between supervised and unsupervised learning and focuses on learning from behavior, experiences, and decisions. The algorithm learns from false suppositions. A case of reinforcement learning includes the video games played by the computer and the computer learns how to play in a considerable scenario. Learning algorithm takes decisions so that outcome it produces would be prescriptive. firstly, when looking at the video, it can be observed that the program is uncoordinated and unskillful, but it constantly improves with training and becomes a champion.

We can choose any of the three learning techniques for training then test it on an unknown dataset. Accurate results will be achieved by splitting data into 75:25 for training and testing.

Features:

For testing the data, preprocessing is always required. Handling large volume of data is difficult so reducing the dimensions of the given dataset into some informative features having the same accuracy and completely well-defined and illustrate the original dataset known as feature extraction. Handled data is stored in the variable and then characteristics are given as input to learning algorithms. Now feature detectors are becoming common so complete datasets can be used in machine learning algorithms.

Generalization:

Generalization is the perception that how well the machine learning model learned the concepts when it is trained with specific examples and gives true results for testing data. For preprocessing learning algorithm must be generalized to handle the unseen data. As the training data does not include all possible inputs so generalization of the algorithm is important.

When the model is too simple that means it has few features known and regularizes data. This is known as under-fitting. It includes all the instances as each instance cannot be included by the training dataset.so the algorithm that learns through the data must be generic in case any unknown example can be tested. If the model is unpretentious then results obtained will be below average. And if a sophisticated model is used then it can badly generalize the data and over-fit it. When the proportion of training data is very low and testing data is high so cause over-fitting. Errors can be determined by the performance of the algorithm. Mean squared error is the loss function to allocate the rate of errors. Generalization is the perception that how well the machine learning model learned the concepts when it is trained with specific examples and gives true results for testing data. For preprocessing learning algorithm must be generalized to handle the unseen data. As the training data does not include all possible inputs so generalization of the algorithm is important.

1.3 Neural Networks

Neural networks were applied for the description of images to improve performance and accuracy. Neural networks were proposed in the early 1950s [9] with the idea of Neuromorphic Learning Computation. The neuron is the principal constituent of

Neural networks. Advanced neural networks have introduced the concept of a perceptron. in this part, we will start with basic how and why neuron is a basic element of neural networks and how can we improve accuracy.

Origin:

Neural networks are most commonly known as Artificial Neural Networks. They were designed to imitate the functioning of the human brain. The human brain can perform complex tasks in parallel with millions of neurons, but the computer can perform them as mathematical equations through consecutive computing. With the advancement in engineering and mathematics research, neural networks are also formulated as a result. Perceptron is weighted down with a different number of neurons. Thus, the perceptron is a building block of Advance neural networks. Perceptron performs rectilinear functioning and produces binary output i.e. 0 or 1. They are associated with weights and biases. Weights and biases are hyper parameters which can be tuned by the user. If the desired output is not achieved, then weights must be updated for the given input vector. Neuromorphic Learning Computation is used to determine the size of weights. constrains of this algorithm is that it can only be applied on the dataset which is uniformly divided.

1.4 Convolution Neural Networks:

Convolution neural networks are used for classification and object detection in images. These deep neural networks are used for facial recognition, human activity recognition, tumor detection, and many other applications related to computer vision and natural language processing. For recognition of objects through machine learning algorithm some problems that may occur are:

- Image segmentation
- Illumination
- Contortion
- Outlook

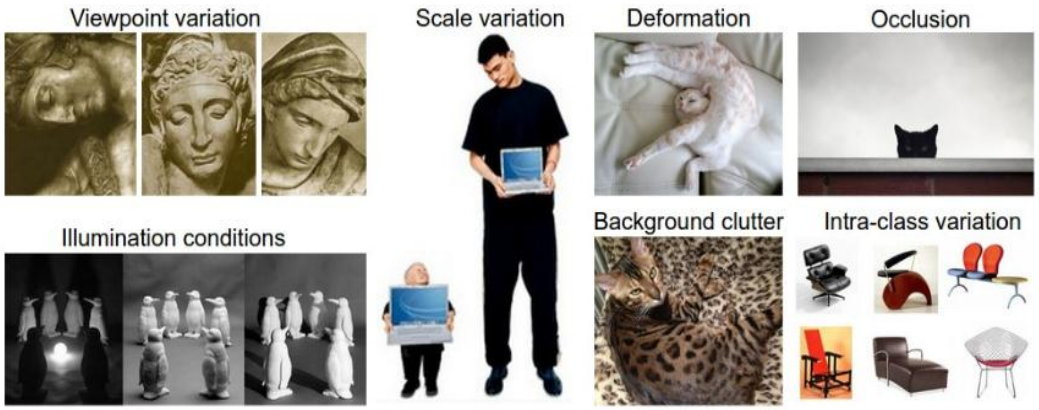


Figure 0.1 Image Classification Inter-class Variations

For the solution to the above-mentioned problems, a convolution neural network is used. For our project images are taken as inputs then classified and localized to perform object detection and then perform segmentation as given in figure 2. This process can be done through the convolutional neural network. Object detection from color images with different dimensionalities, different pixels scattering, classification, and multiple objects is a difficult task. Convolution Neural network consists of a stack of layers in which each layer performs specific functionality. Structure of ConvNets mainly includes three layers i.e. convolution layer, pooling layer, and fully connected layer. Description of each layer is given below:

- Convolution layer: This layer is the very basic and important level in the structure. The process of computation starts from here. This layer contains filters that learn from the environment and are hyper parameters whose width and height are determined by the programmer. A filter is also known as a kernel or receptive field. For the processing of images, we use the filters of the minimum size I.e.3x3x3 where 3x3 represents the height and width of pixels and 3 represents the channel i.e. RGB. Throughout the whole transmission, the input image is convolved with kernel starting from the top left of the image till the right bottom of the image pixel it computes the dot product and of each pixel with the kernel. Then with the size of the kernel, it computes the mean value and updates the central pixel value at every spatial location. Different filters are applied such as smoothing, sober filter, etc. but

exactly filters are not circumscribed as each filter learns from the training of images dataset. Before filtering, we also do zero-padding around the image.

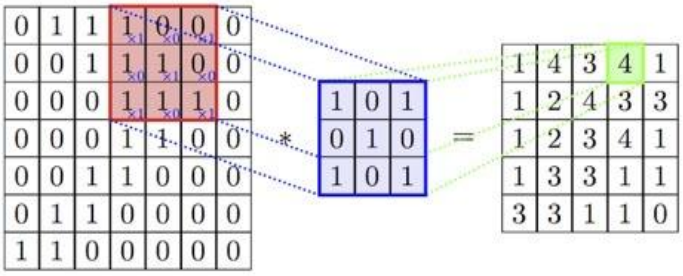


Figure 0.2 Working of image matrices with the kernel in CNN

- Pooling layer: To decrease the number of hyper parameters and computation power we use the pooling layer after the convolution layer thus the spatial dimensions may also reduce. This layer is used to avoid over-fitting. In this layer, down sampling is performed with a stride of 2 to resize the image using MaxPool operation in spatial dimensions.
- Fully connected layer: In this layer, the number of neurons is connected to all other neurons in the network. Activation function will be performed with input matrix multiplication and predicts the output just like Neural Networks.
- Relu: Rectified linear unit activation function is used as it is much faster, and this increases the non-linearity. Image pixel values will be non-linear in nature s this function is used to add non-linearity.

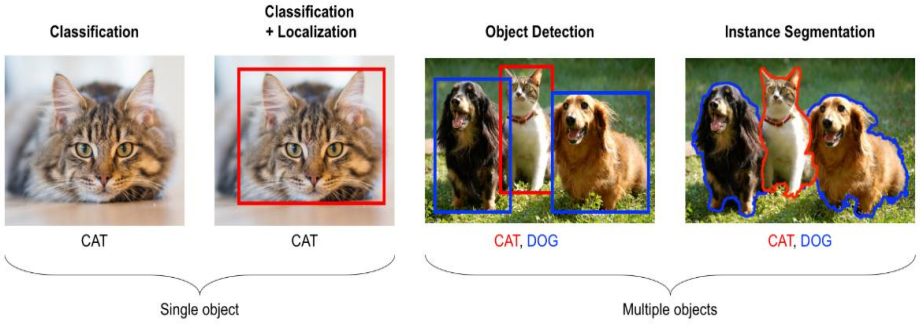


Figure 0.3 Object Detection and Classification using CNN

1.5 Deep Learning

After the origination of neural networks, deep neural networks were proposed in 1998 [10]. Back propagation algorithm was then introduced as an optimization algorithm for the improvement of cost function. Deep neural networks are too much complex for training and because of this complexity they oblige modest framework for hardware and software. Deep learning is subset of Artificial Intelligence. It is referred to imitate how human learns through different experimentation and knowledge. It is used for prophetic scrutiny.

Machine learning approaches consists of 1 dimensional problem, but deep learning approaches are piled in pecking order with high generalization and complexity. This concept can be figure out by the example when small child starts learning the language first that kid learns the objects like car or cat by noticing them then by observing he learn the characteristics of different objects with time. Thus, with in different levels he learns the attributes of specific object and able to differentiate between different objects.

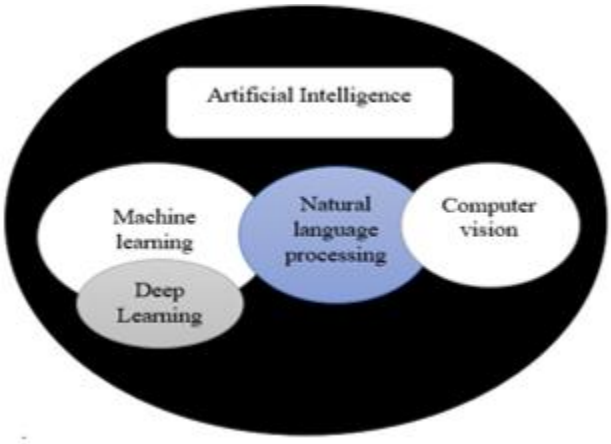


Figure 0.4 Relationship of subdomains of AI

1.6 Problem Statement:

We are observing a predominant use of surveillance cameras for both indoor and outdoor environments for different angles which includes security, tracking of different people, and traffic monitoring. The main problem in the surveillance system is identifying, verifying individuals over various cameras or multiple cameras. If a person is seen in a camera view also seen in another camera. The problems that researchers faced have created great attention in the Computer Vision field, the desired level of accuracy through person tracking by face recognition couldn't be achieved so far, still, several problems exist to hold up significant improvement in the prediction performance. First, it is difficult to analyze spatial continuity between two non-overlapping camera views particularly when the camera views don't overlap. Second, lighting conditions may vary in different views. For example, an indoor environment has different illumination condition from the outdoor environment, which make it difficult to adjust [10]. Third, is changing in pose and occlusion. For example, if the image of a person taken from one side by another camera may not be identical, it can be taken from the front view. The similarity between the same dressings also makes problems. Many people wear jeans, or white and black are fairly common in clothing. After all, present processes suffer from the poor quality of the input images. Because cameras are usually arranged to monitor large-format images of individuals with low resolution. Therefore, fixed visual and scale models are required to solve the problem. In this project, we will do the tracking of the person using different cameras or networks of cameras and it will also give the current location and current time of the person. This project will provide security to keep tracking of the person which cannot be possible through the ordinary camera. Earlier work consists of the still image but in this project, we will do the person detection and Tracking in a real-time video which is a challenging task in the field of computer vision.

1.7 Objectives

The goal of our project "Real-time face detection and person tracking" is to provide a useful tool for surveillance and to track detected persons within a video sequence to generate person's path, location, and time of person for security to monitor people in public places. The main objectives are as following,

- To provide the tracking of a specific person.
- To Track the same person over a network of cameras or different cameras.
- To provide a useful tool for video surveillance.
- To recognize an individual who was previously observed over a camera network.
- To provide the present location and time of the specific person who was detected by the camera.

1.8 Scope

We hope that our project can be a productive and effective application for the security purpose to monitor people in public places and track a person's location and time. Our product will provide security to keep tracking of the person which cannot be possible through the ordinary camera. Our product gives more accurate results and effective tracking than other security systems products.

CHAPTER 2

2 LITERATURE REVIEW

2.1 Face Detection:

Face detection is a computer technology used in a variety of applications and identifies human faces from digital images. Face recognition also refers to the psychological method by which people locate and deal with the faces in a visual scene.

- Feature based method.
- Appearance based method.
- Knowledge based method.
- Template matching method.

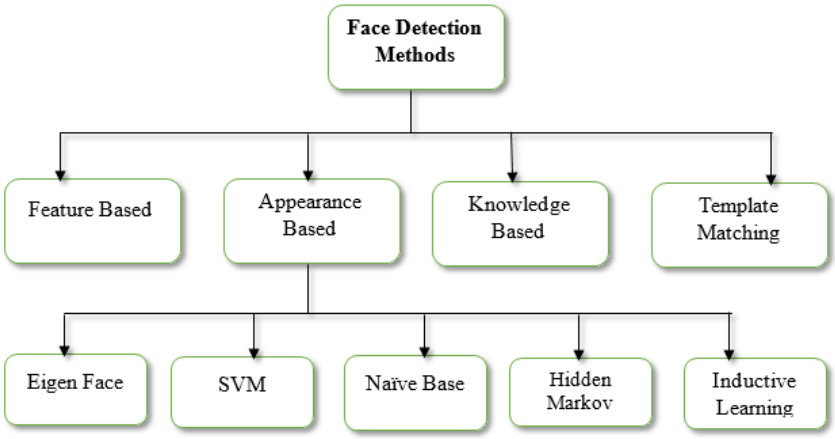


Figure 0.1 Face Detection Techniques

2.1.1 Feature Based Method

Objects are generally recognized by their characteristic or unique properties. There are various features of the human face that can be recognized between the face and many other objects. It locates faces by extracting structural features like eyes, mouth and nose, etc. and then using them to recognize or detect a face. In general, some kind of qualified arithmetical classifier is then useful to separate facial and non-facial regions. Numerous methods for extracting properties have been proposed in the literature. The

problem with these algorithms is that these functions are corrupted due to lighting, occlusion, and noise.

Human faces have specific textures that can tell the difference between a face and other objects [11]. In addition, the functional border can help to recognize objects on a face. The feature-based technique is used to trace faces by mining the structural features of a face. It is first formed as a classifier and then used to distinguish between facial and non-facial regions. The idea is to go beyond the limits of our instinctive facial knowledge. This approach, which is divided into several phases and even photos with multiple faces, shows a 94% of success rate.

2.1.2 Knowledge Based Method

The knowledge-based method depends on rules and is usually based on the human knowledge to detect or recognize faces [12]. For example, a face must have a nose, eyes and a mouth at certain intervals and positions relative to each other. The big problem with these methods is that, it has difficulty in creating a suitable set of rules. There could be many false positives if the rules were too general or too detailed. This approach alone is not enough and cannot find many faces in several pictures.

2.1.3 Template Matching Method

The template matching process uses predefined or configured face models to locate or recognize faces through the correlation between models and the input images [13]. For example, a human face can be divided into eyes, facial contour, nose and mouth. In addition, a face model can be easily created by edges using the edge detection method. This approach is easy to implement, but not suitable for recognizing faces. However, deformable models have been proposed to solve these problems.

2.1.4 Appearance Based Method

The appearance-based method is basically depending on a number of facial image training provide to determine face models. The appearance-based method uses statistical analysis and machine learning techniques to determine the relevant properties of facial images [14]. This method is also used in the extraction of features for facial recognition.

The appearance-based model is divided into the following sub-methods for facial detection which are as follows:

- Eigen face-based method.
- Support vector machine (SVM) method.
- Sparse network of winnows method.
- Naïve base classifiers.
- Hidden Markov model.
- Information theoretical approach.
- Inductive learning method.

2.2 Face Recognition

Face recognition is a technology with which a person can be identified or checked using a digital image or a video frame from the video source. Due to its many practical applications in the fields of information security, biometric, and smart cards. Face recognition is one of the researches for recognizing surface shapes and fractal features. Which is used in Law enforcement, access control and security system. Face recognition methods are:

- Holistic matching method.
- Feature-based or structural method.
- Hybrid method.

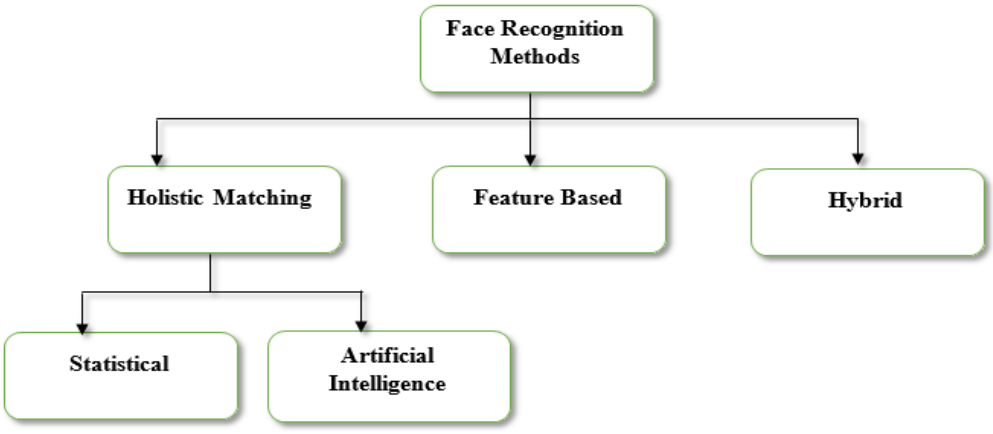


Figure 0.2 Face recognition techniques

2.2.1 Holistic Approach Method

With holistic approach, the entire face region is considered as input for the face detection system. It compares the similarities of the entire face and ignores individual features such as eyes, mouth, nose, etc. [15].

Holistic matching methods are categorized into two parts:

- Statistical methods.
- Artificial intelligence (AI).

Best examples of holistic methods are:

- Eigen faces (principal component analysis PCA).
- Fisher face (Linear discriminant analysis LDA).
- Independent component analysis (ICA).

2.2.2 Feature Based Matching

In contrast to a holistic, feature-based approach should consider the individual features of the face such as eyes, nose, mouth, birthmarks and ears and compare the similarity between the images [16]. Another approach in the field of facial recognition involves facial recognition by recognizing hexagonal lines. This approach is based on edge detection for recognizing and recognizing faces with hexagonal facial lines. Face recognition by using heuristic parameters and storing them in the database before the search can be analyzed. This approach focused mainly on the nasal part of the captured images, followed by the conversion of the grayscale and the transformation of the intensity.

2.2.3 Hybrid Face Recognition Method

Hybrid facial recognition systems use a combination of holistic extraction methods and features base method. In general, 3D images are used in hybrid methods. The image of a person's face is recorded in 3D, which allows the system to note, for example: the orbital curves of the eye or the shape of the chin or forehead. Even a profile surface would be useful because the system uses depth and a measurement axis that provides enough information to create a complete face [17].

For face detection specifically there are two main pre trained classifier.

- Haar cascade classifier.
- LBP cascade classifier.

2.3 Person Tracking

Human tracking is the process of temporally assigning human detections within the video sequence to create persistent paths or trajectories of the people. Detection and tracking of people are generally considered to be the first two processes in a video surveillance pipeline and can support higher-level reasoning modules such as detection, action and dynamic scene analysis.

2.3.1 Previous work:

Previous work of person tracking using features of human, based on the deep sort algorithm. Deep sort algorithm is basically used for the object tracking and a person tracking.

2.3.2 Deep Sort:

The most popular and one of the most widely used sophisticated object tracking framework is Deep SORT, an extension to SORT "Simple Real time Tracker" [18].

2.3.3 Our approach:

In our project we will use the time stamp method for tracking. It is basically creating the txt file and update it when the person detects or recognize successfully. It will update the file based on name of person, location of person, time of passing from the camera, and the camera number.

This file keeps the record of a person. So, whenever we want to check the record of the person or to track the person, we can easily get the data from this file.

Table 2.1 Summary of Earlier Techniques Used

References	**Title**	**Author Name**	**Model**	**Technique Used**
Shishira R Maiya 2016 (Yang et al., 2016)	Deep sort: deep learning to track objects in a video	Shishira R Maiya	Machine Learning	Deep sort algorithm
Nicolai woke 2016 (Mathews et al., 2015)	Simple online and real-time tracking with a deep association metric	Nicolai woke Alex Bewley Dietrich Paulus	Recurrent Neural Network	Deep Association Matric, Deep appearance descriptor.
Esther Koller-Meier 2015 (Xu et al., 2015)	Online Multi-Person Tracking-by-Detection from a Single, Uncallibrated Camera	Michael Breitenstein Esther Koller-Meier Luc Van Gool	Supervised Learning	Markov model, boosted classifiers, Detector confidence particle filter.
Swapnil v. Tathe Sandipan p. 2017 (Shetty et al., 2017)	Real-time human detection and tracking	Swapnil v. Tathe Sandipan p. Narote	Machine Learning	Face detection using skin color-based method, color model to detect skin regions, Projection function and pixel count for eye region detection.
Gerais Khan 2017 (Ren et al., 2017)	Multi-Person Tracking Based on Faster R-CNN and Deep Appearance Features	Gerais Khan Zeeshan Tariq Muhammad Usman Ghani Khan	Faster R-CNN	Faster R-CNN for human detection, Kalman filter, CNN for appearance features

CHAPTER 3

3 DESIGN AND METHODOLOGY

For Real-time Person tracking, the first stage is to detect the person using a single camera or multiple cameras. For face detection, first, we have to preprocess the images. For image processing by converting RGB images into grayscale, and for enhancing image we use contrast stretching. Pre-process images to remove noise and blurriness in images and then we detect the person face in an image. After face detection, we trained our model to extract facial features and stored in the database, and then apply Haar cascade classifier for face Recognition. For Person tracking, We introduced location and time stamps to assist the person tracking task which will show us the time when the person has passed from the camera and also tells us the current location of a person and camera number.

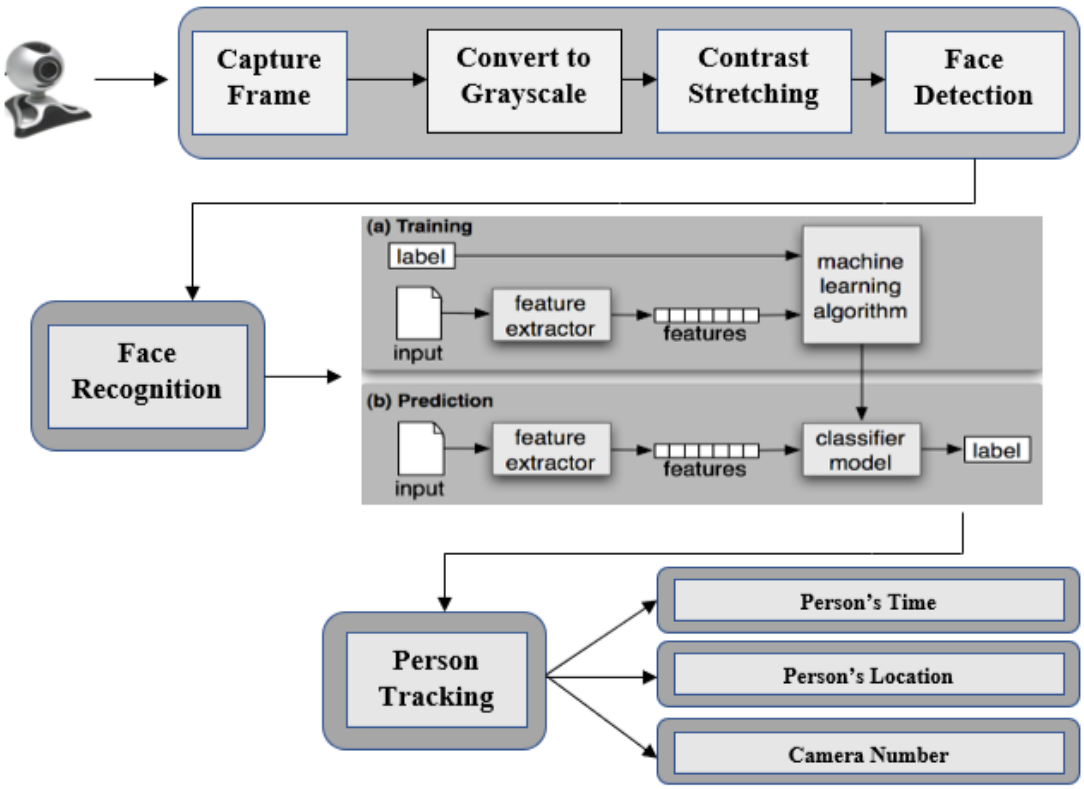


Figure 0.1 Block Diagram

3.1 Images Preprocessing

The first step is the preprocessing of images from the dataset. For face detection, the first step is to preprocess the images by converting RGB images into grayscale, and for enhancing image we use contrast stretching. Pre-process images to remove noise and blurriness in images and then we detect the face in each image.

3.1.1 Grayscale Image

First, pre-processing method for face detection is changes the RGB into grayscale. As the image is in RGB format, so it means that we have added R+G+B then divide it by 3 to get a grayscale image. Conversion from real image into Grayscale to reduce the complexity, from 3D pixel value (R, G, B) into Grayscale 1D.The formula of convertion from real image to gray scale is given below:

$$\text{Grayscale} = \alpha * RED + \beta * GREEN + \gamma * BLUE \,/\, 3$$

3.1.2 Contrast Stretching

Contrast stretching is applied to images so that images can be seen clean. Contrast Stretching sometimes is also called normalization. The formula used for contrast stretching is given below:

$$Io = (I_i - Min_i) * (Max_o - Min_o / (Max_o - Min_o + Min_o)$$

Here, Io is the output pixel, Ii is an input of pixel of an image, Mini is the minimum pixels of an input image, Maxi is the maximum pixel of an input image, Mino is the maximum pixel of an output image and Maxo is the maximum pixel of an output image. After using this we got a new image where the pixel intensity is higher than the dark image's pixel intensity.

3.2 Import Modules

Modules that are needed to perform face detection are os, cv2, and NumPy.Cv2 is an OpenCV module that contains a function for face detection and recognition. OS library will provide a function that interacts with the operating system.

3.3 Face Detection

Face detection is an initial step of facial recognition and faces identification. Face detection determines the location and the size of a person's face in digital images by ignoring other images.

For face detection, we preprocess the images by converting RGB images into grayscale, and for enhancing image we use contrast stretching. Pre-process images to remove noise and blurriness in images and then we detect the face. After face detection, once we find the region of interest which contains face images, then we use it for training and recognizer. For face detection, we have used the Haar cascade classifier, which is provided by OpenCV.Haar cascade face detected XML file used for detecting a face from images. Cascade classifier is loaded by using Cv2, Cascade Classifier takes the path to cascade .xml file, and detect the face.

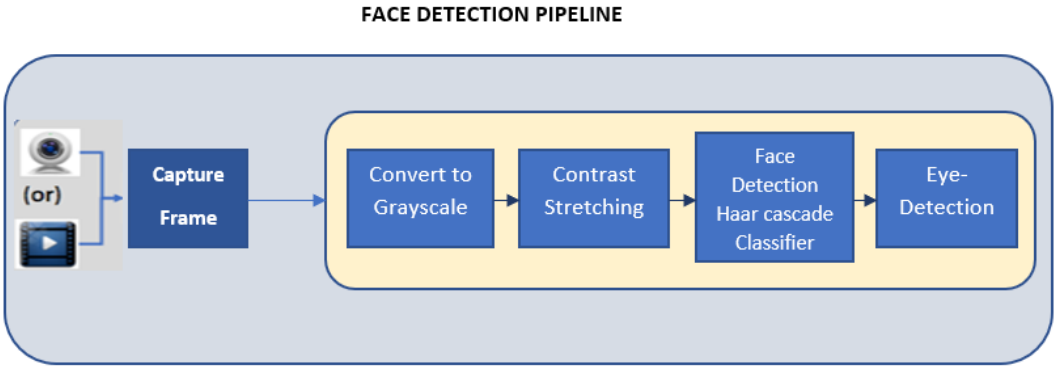


Figure 0.2 Face Detection Pipeline

3.3.1 Haar Cascade

Haar features are wavelet-based features that decompose an image. Wavelet Haar is a rectangle shape wave that has two intervals, one is high and one is low. For the two dimensions, it has two sides, One side is light and the other side is dark. The cascade classifier has the function to combine many features effectively and more efficiently. At the start,image consist of real RGB value and then refine an image in rectangular shpae. Each shape is refined then a limited threshold has occurred which shows some light and the dark area. Haar features extracted by averaging the resultant value above a certain threshold.

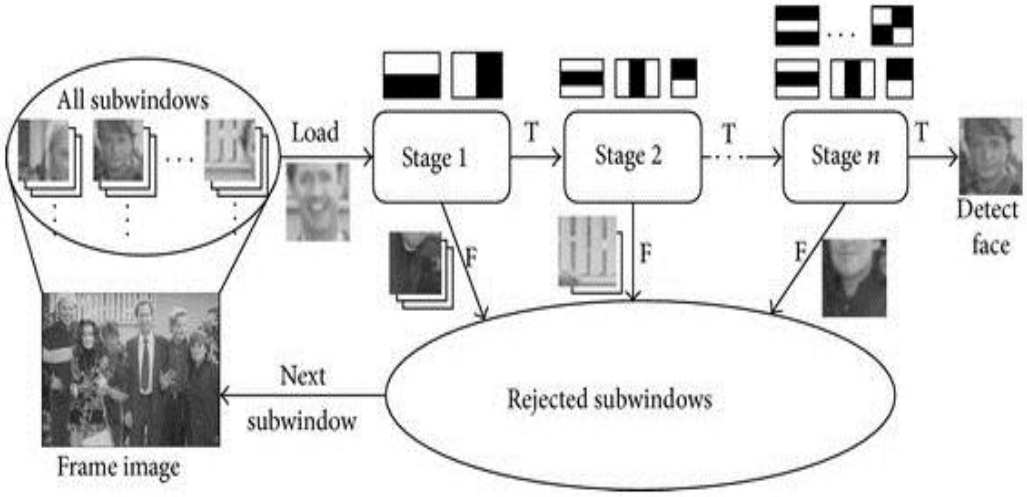


Figure 0.3 Architecture of Haar Cascade Classifier

In general, every process in the Haar-like feature needs to get every single pixel in an area and then passed by applying integral images, which contains sum of all gray values.Therefore, it only calculates the four pixels value looks up from the integrated image.

Haar-like features are deploy ML, which is a cascade classfier trained function with some human face and without human face images and upgrades their classifier to locates the object within some different images. The classifier trains its classifier with some positive and negative images. Hence, the classifier can detect images from surrounding objects. For Example, if size of image is 24x24 window, there are 160000 features needed to verify this method. For the computation of every single-feature, total pixels in an image are required under a white and black rectangle.

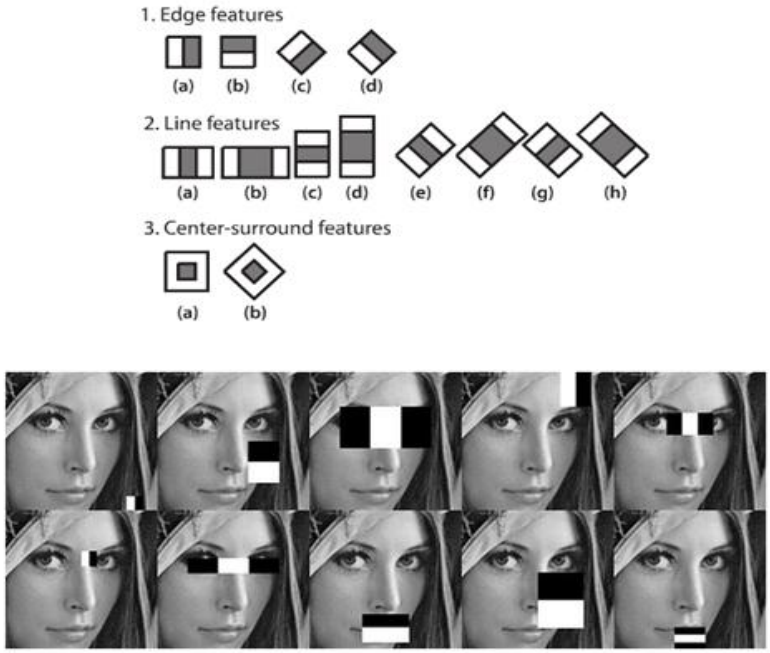


Figure 0.4 Face Detection using Haar Cascade Classifier

$$\text{Feature} = \sum ie\{1....N\} wi.\text{Rec}Sum(x,y,w,h)$$

Recsum (x,y,w,h) the sum of all pixel values in an enclosed rectangle, whereas x,y,w,h are the coordinates, rotation, and the dimensions of that rectangle. Haar wavelet classifier that is used to extract features from the face.

Each x and y location of the integrated image is the sum of all pixels values above and left of x and y location.

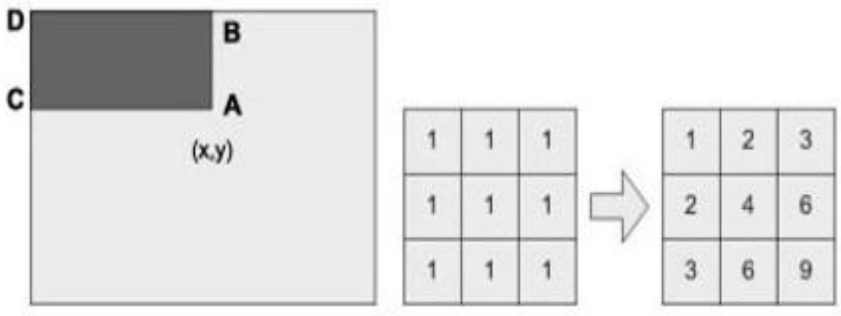


Figure 0.5 Integral Image

3.4 Camera Connectivity

Figure 0.6 Hercules Duplex HD webcam

Our project-based on person detection and tracking using multiple webcam's.In our project, we have used two Hercules duplex HD webcam's, that are used for face detection and person tracking for surveillance. We install a new Xtra Controller+ driver software pack of webcam. Then connect the camera with PC.

3.5 Dataset

we have created our dataset in which each person has 500 images with true labels. we have created the dataset of images. Initially, we have gathered all the images of faces. First, we need the data which can be used for training. We took a lot of face images with different backgrounds, different faces, and different lighting conditions. All images are so big because they have high resolution so we transform them into a lower scale so that during training the processing will be faster.

Figure 0.7 Face Images Dataset

3.5.1 Packages Installation

Conda Install -c Anaconda Protobuf

Pip Install Pillow

Pip Install Lxml

Pip Install Jupyter

Pip Install Matplotlib

Pip Install Numpy

Pip Install OpenCV-Python

Pip Install Pandas

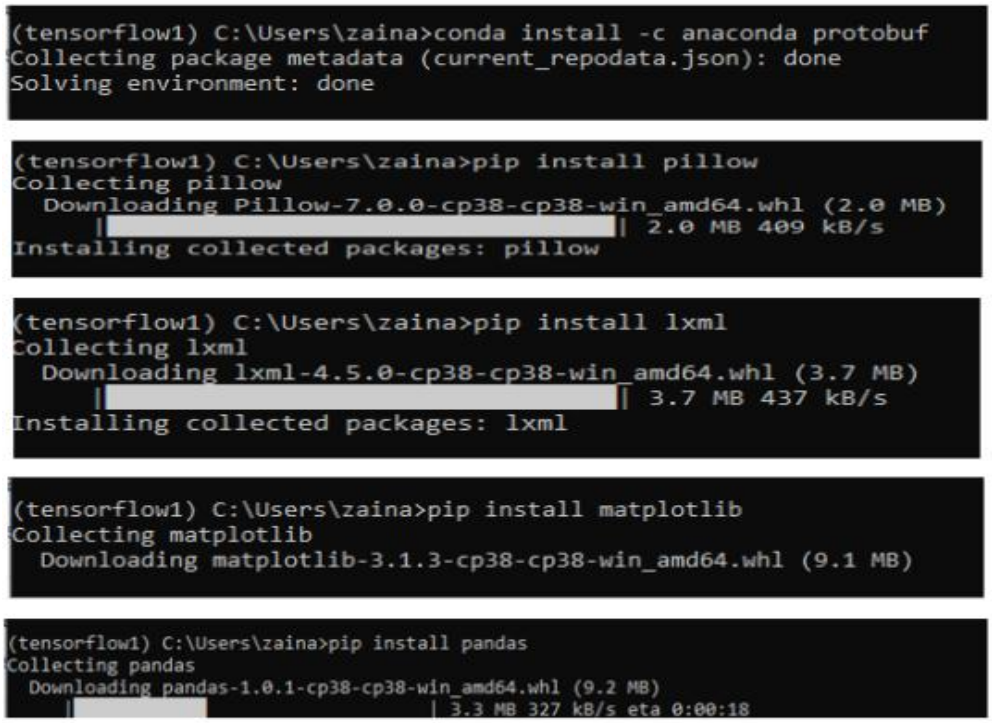


Figure 0.8 Packages Installation

3.5.2 Labeling Dataset

To label the data we have used the labeling tool. we have labeled all the faces, save them into .xml file and convert all XML files into CSV files. The images folder contains the trained images. The CSV file present in test set is different from the training set. Test file have all the test images, but they don't have not labels.

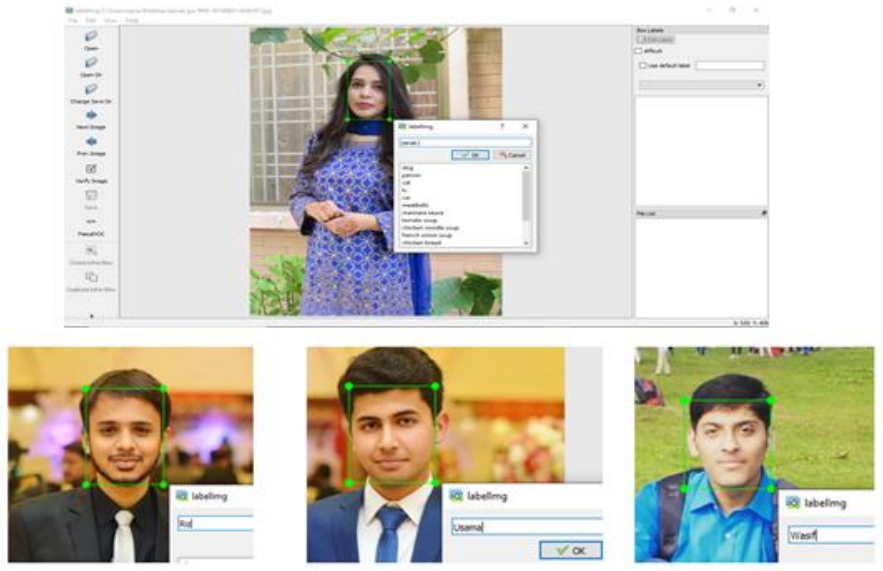


Figure 0.9 Labelled Face Images

3.6 Face Recognition

After face detection, the next step is face recognition. Face recognition identifies person seen on camera. We trained our dataset using Haar cascade classifier, after training it to generate .yml file which is having our extracted facial features. When a person passes from camera number 1, they compare selected facial features from a given image with faces within extracted features in the database and recognize that person with the given label.

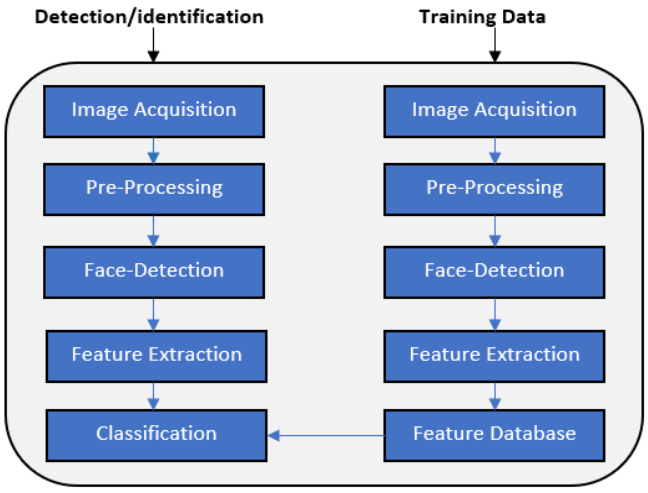


Figure 0.10 Face Recognition Pipeline

3.7 Person Tracking

Face detection is the initial step of face recognition and the part of the face identification. After face detection, we have used the Haar cascade classifier for face recognition. We introduced location and time stamps to assist the person re-identification task which will show us the time when the person has passed from the camera and also tells us the current location of the person.

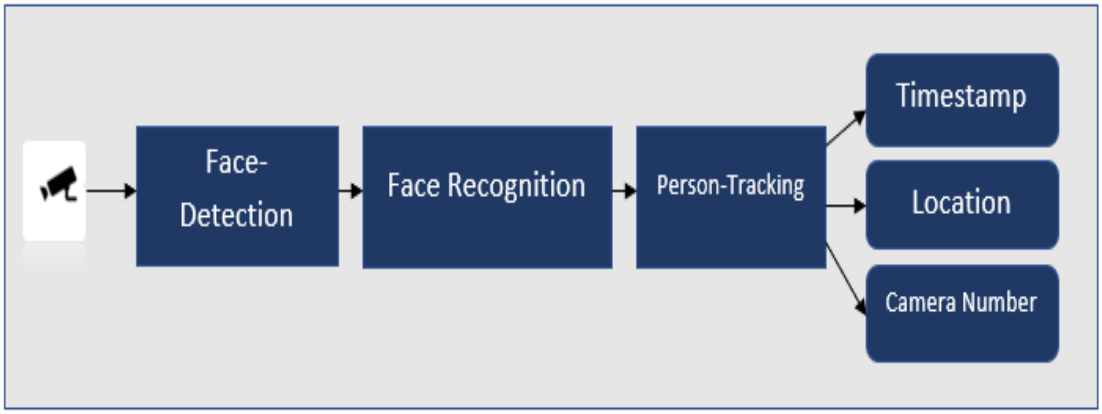


Figure 0.11 Person Tracking Block Diagram

Specify TIMESTAMP to sort the elements through time of a person's path. To track the person first. we have information of the detected person, the person detection array hold all the information about the detected person. For Example, the time of the person was detected (Timestamp), and the position of the person in the video frame at the time they were detected. When a person will pass from camera number 1 it will recognize the person and store the time. Whenever the person will cross the camera it will store the time and also tells us that the person has passed from camera 1 or camera 2 etc. It will store the record of a person's activity. So, it would be helpful for the tracking of a person.

CHAPTER 4

4 RESULTS AND OBSERVATIONS

First, our system obtains the capture frame form web cam. Frame is basically a picture. After that this RGB frame is converted into grayscale to decrease intensity level which is useful for the edge detection process in Haar cascade classifier. After converting the frame into grayscale, we perform contrast stretching to enhance the image which is basically normalization of an image. When the frame is done with preprocessing then it is passes through the face detection process. In face detection, Haar cascade classifier is used to detect the face based on facial features like nose, eyes etc. basically edge detection process is used in Haar cascade classifier. After the face detection process face recognition is implemented.

Then, we have created our dataset in which each person has 500 images with true labels. We have created the dataset of images. Initially, we have gathered all the images of faces. First, we need the data which can be used for training. We took a lot of face images with different backgrounds, different faces, and different lighting conditions. All images are so big because they have high resolution so we transform them into a lower scale so that during training the processing will be faster. After training we obtain accuracy 85.64%. Then we apply Haar cascade classifier for face recognition.

When the face recognition is done then the person tracking can be possible. In our project we will use the time stamp method for tracking. It creates the txt file and update it when the person detects and recognize successfully. It will update the file based on name of person, location of person, time of passing from the camera, and the camera number.

4.1 Image Acquisition

Initially we acquire image and display it.

Figure 0.1 Image Acquisition

4.1.1 Conversion of RGB Image to Grayscale

For face detection, the first step is to preprocess the images by converting RGB images into grayscale, and for enhancing image we use contrast stretching. Pre-process images to remove noise and blurriness in images and then we detect the face in each image.

Figure 0.2 Conversion from RGB Image to Grayscale

4.1.2 Normalization

After converting the frame into grayscale, we perform contrast stretching to enhance the image which is basically normalization of an image. When the frame is done with preprocessing then it is passes through the face detection process.

Figure 0.3 Normalized Image

4.2 Face Detection

In face detection, Haar cascade classifier is used to detect the face on the basis of facial features like nose, eyes etc. basically edge detection process is used in Haar cascade classifier. It creates a boundary box around face for detection.

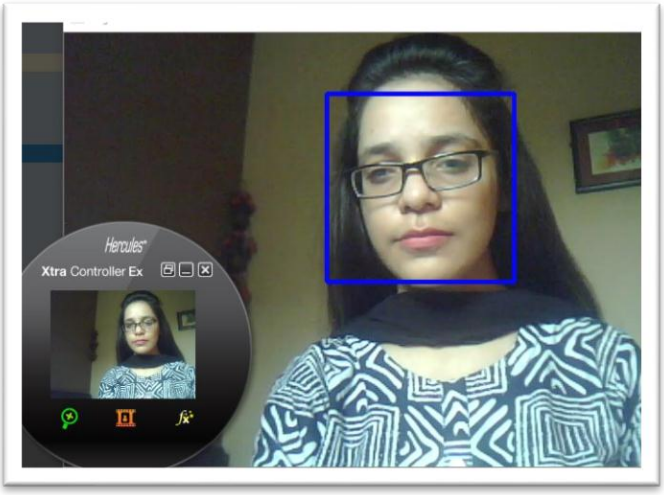


Figure 0.4 Real-Time Face Detection

4.2.1 Eyes Detection

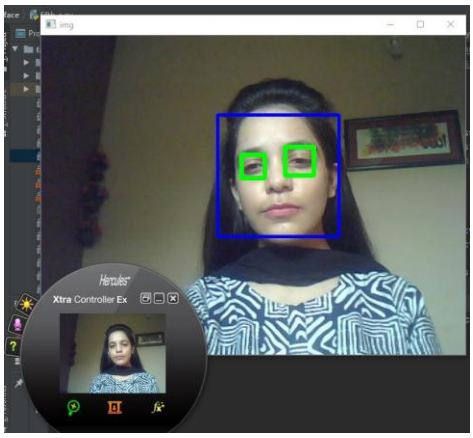


Figure 0.5 Real-Time Eyes Detection

4.3 Face Recognition

For face recognition we used LBPH. For LBPH first of all we have to train the algorithm by using dataset which contains the facial images of the people we want to recognize.

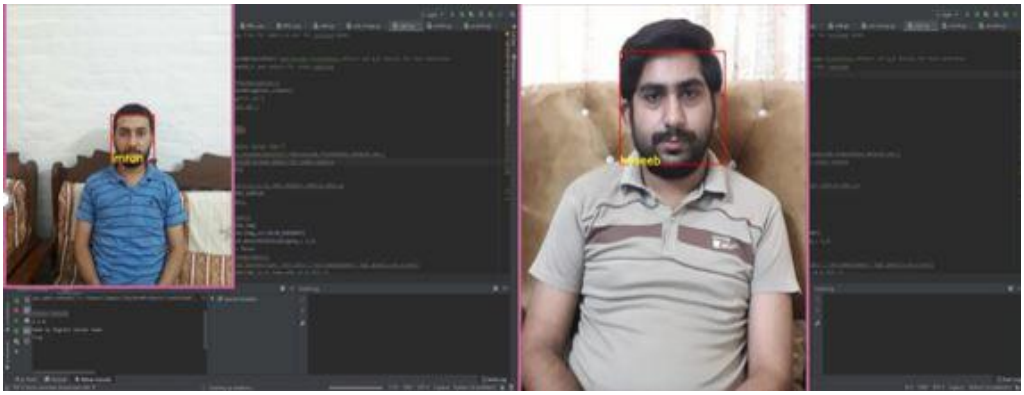


Figure 4.6 Face Recognition

4.4 Person tracking using single camera

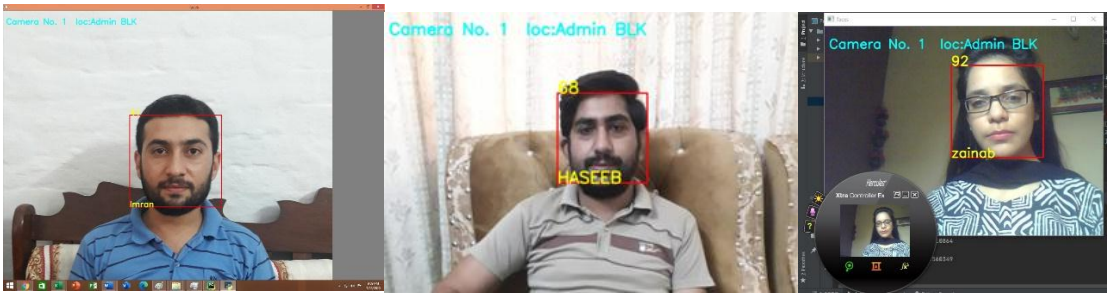


Figure 0.7 Person Tracking Using Single Camera

```
2020-07-18 21:09:14.817051 Camera No. 1  loc:Admin BLK Imran
2020-07-18 21:11:41.366549 Camera No. 1  loc:Admin BLK Haseeb
2020-07-18 21:11:41.786471 Camera No. 1  loc:Admin BLK Zainab
```

Figure 0.8 Tracking Person Time

4.5 Person tracking using multiple cameras

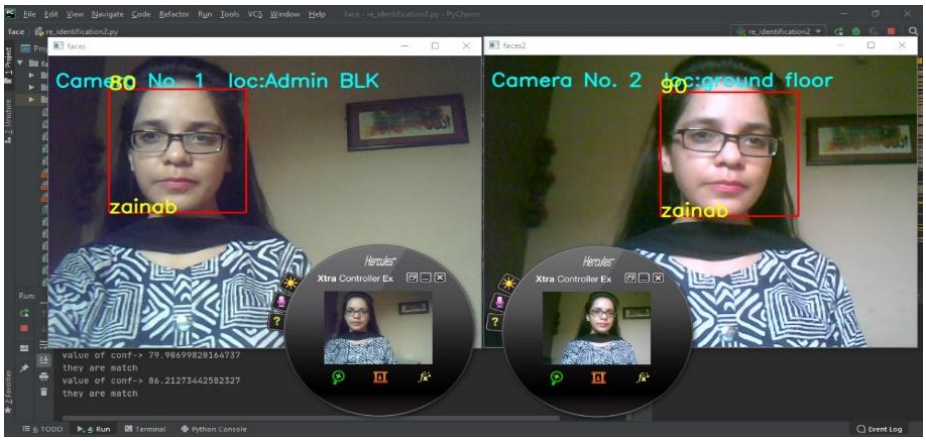


Figure 0.6 Person tracking using multiple cameras

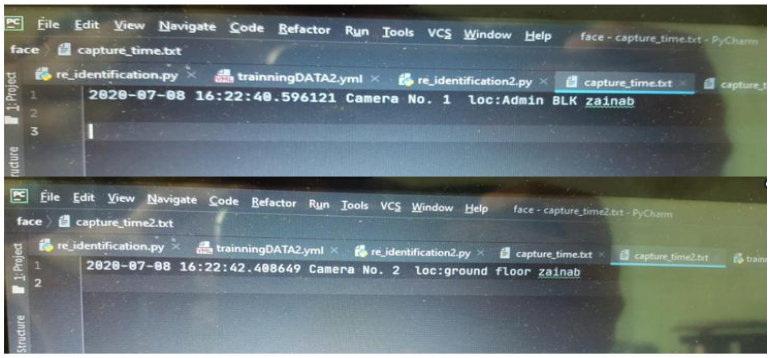


Figure 0.7 Tracking person's time using multiple cameras

4.5.1 Intruder Detection

Figure 0.8 Intruder Detection

4.6 Accuracy

Table 4.1 Accuracy of recognition and tracking

Names	Recognition accuracy	Tracking Accuracy
Muhammad Imran	92	84
Muhammad Haseeb	92	84
Zainab Iftikhar	92	84

The accuracy of confusion matrix is calculated by the given formula

$$\text{Accuracy} = \frac{true\ Positive + true\ negative}{true\ positive + false\ positive + true\ negative + false\ negative}$$

4.7 Negative result

As it is clear from the figure given below that the label assigned to the person is false this is because of less amount of pictures in dataset so it will cause the error in recognition.

Figure 4.12 False output

4.8 System Implementation

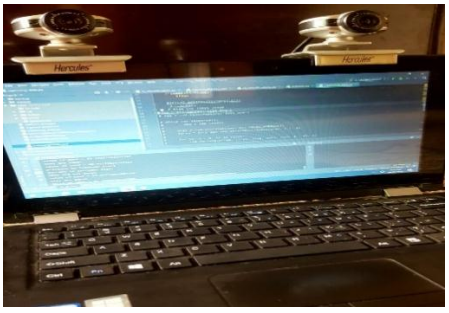


Figure 0.13 System Implementation

CHAPTER 5

5 CONCLUSION

We must perform Real-Time person detection and Tracking for Surveillance. Person detection and tracking used for locating and tracking the person in the video stream. Person detection locating all persons present in an image, at all possible scales. Person Tracking is allocating the person detected within a video sequence to generate person paths, location, and time of the people. Nowadays security systems are used extensively for monitoring purposes to detect person in different locations.

For Real-time Person tracking, we detect the person using a single camera or multiple cameras. For face detection, first, we must preprocess the images After face detection, we trained our model to extract facial features and stored in the database, and then apply LBPH for face Recognition.We have used supervised learning method to train our model. For Person tracking, We introduced location and timestamp to track person. From the results, it is concluded that if we use a large dataset i.e. Neu-Gou/awesome-person tracking dataset we can improve our results. Our project is implemented through a supervised learning approach. We have used a Machine learning model for person detection and tracking. In the future, this work can be enhanced in Deep Learning Models I.e. Convolutional Neural Network (CNN), Region-based Convolutional Neural Networks (R-CNN), Faster-RCNN, or YOLO based techniques for face detection and Person Tracking. These techniques can help us with the enhancements of the presented work.

CHAPTER 6

6 FUTURE WORK

Real time face detection and person tracking is design to provide high security systems to monitor people in public places and will provide security to keep tracking of the person which cannot be possible through the ordinary camera. The system designed can be intensely improved and we will discuss some of the improvements below.

we will extend this existing system by embedding multiple cameras at different locations and extend the database by adding a lot of people's images. We will improve the performance at tracking challages involving pose variation, occlusion and rotation. As lighting conditions may differ in different views. For example, an indoor environment has different illumination condition from the outdoor environment, which make it difficult to adjust, so we can extend this work by using high resolution cameras which enhance the brightness so that we can use it for both indoor and outdoor environment. We are looking forward to enhance its acuraacy and efficency by using large dataset. Moreover, this work will be extended to track the the person by leading towords Deep learning. Deep learning will handle large dataset and also improve the system acuracy. We have used a supervised Machine learning model for person detection and tracking. In the future, this work can be enhanced in Deep Learning Models I.e. Convolutional Neural Network (CNN), Region-based Convolutional Neural Networks (R-CNN), Faster-RCNN, or YOLO based techniques for face detection and Person Tracking. These techniques can help us with the enhancements of the presented work.

REFRENCES

[1] Md. A.R. Ahad, “Computer vision and action recognition: A guide for image processing and computer vision community for action understanding,,” 2016.

[2] P. Mahdi and Mahmudul Habib, “"Face recognition-based real-time system for surveillance",” 2016.

[3] X. p. A. Milella, ““Real-Time Human Objects Tracking for Smart Surveillance at the Edge". IEEE International Conference on Communications, 2018-May,p. 1–6.,” (2018). .

[4] D. Baikova, “ “Real Time Multiple Camera Person Detection and Tracking." October, p. 155,,” (2018)..

[5] Khan, M., Chakraborty, S., Astya, R., & Khepra, S., “Face Detection and Recognition Using OpenCV.Proceedings - 2019 International Conference on Computing, Communication, and Intelligent Systems,,” 2019.

[6] D. Baikova, “Real Time Multiple Camera Person Detection and Tracking. October,,” p. 155, (2018).

[7] M. B. Chrysos, ““A Comprehensive Performance Evaluation of Face Tracking In-the-Wild.”International Journal of Computer Vision,,” pp. 126(2–4),, (2018). .

[8] P. Viola and M.J. Jones, “"Robust real time face detection, Inter- 725 national Journal of Computer Vision",” p. 137–154, 2018.

[9] M. Papert, “Review of 'Perceptrons: An Introduction to Computational Geometry',” 2018.

[10] R. Rangaswami, “The surveillance system, IEEE Conference on Multimedia and Expo,” 2017.

[11] R. Cipolla, “Feature-Based Human Face Detection,” England, 2018.

[12] S. Abbas Z. Kouzani, “Commonsense knowledge-based face detection,” Budapest, Hungary, Hungary.

[13] Scott Tan Yeh Ping, “FACE DETECTION THROUGH TEMPLATE MATCHING”.

[14] C. Molder and R. Oancea, “Appearance-based facial detection for recognition,” Bucharest, Romania, 2012.

[15] Z. Sasan Karamizadeh, “An Overview of Holistic Face Recognition,” Malaysia, 2015.

[16] S. Manjunath, “A feature based approach to face recognition,” Champaign, IL, USA.

[17] M. Trupti M. Kodinariya, “Hybrid Approach for Face Recognition,” Rajkot, Gujarat, India.

[18] S. R. maiya, “Deep sort algorithm for tracking,” 2016.

[19] A. Habeeb, “"Introduction to Artifical Intelligence",” September 2017.

[20] M. K. Naveen kumar, “A robust method for face recognition and face emotion detection system using support vector machines.2016 International Conference on Electrical, Electronics, Communication, Computer and Optimization Techniques, 2016,,” (2017).

[21] A. Kay, “"Artificial Neural Networks"”.

[22] L. Sirovich and M. Kirby, “ “An approach to face detection and recognition", IEEE International Conference on Recent Advances and Innovations in Engineering,,” 2016.

[23] D. &. S. R. Meena, “An approach to face detection and recognition. 2016 International Conference on Recent Advances and Innovations in Engineering,2016,,” 2016.

[24] R. Storey, “ “Integrated Deep Model for Face Detection and Landmark Localization From"*In The Wild” Images. IEEE Access,* p. 74442–74452, 2018.

[25] R. Rangaswami, Z. Dimitrijevic, E. Kakligian, and R. Chang,, “The video surveillance system" *IEEE Conference on Multimedia and Expo,* 2018.

[26] Zenghui Wang, “"Deep Convolutional Neural Networks for Image Classification: A Comprehensive Review",June-2017.

[27] M. B. Chrysos, ““A Comprehensive Performance Evaluation of Face Tracking In-the-Wild.” *IEEE International Journal of Computer Vision,*” pp. 126(2–4),, (2018)..

[28] Milella, “Real time human object tracking for smart survelliance at the edge,” 2018.