Heaven's Light is Our Guide

DEPARTMENT OF COMPUTER SCIENCE & ENGINEERING

Rajshahi University of Engineering & Technology, Bangladesh

Face Recognition using Viola-Jones Face Detector & Principal Component Analysis Approach

Author

Md. Rezaul Karim

Roll No. : 133097

Department of Computer Science & Engineering

Rajshahi University of Engineering & Technology

Supervised by

Professor Dr. Md. Rabiul Islam

Head

Department of Computer Science & Engineering

Rajshahi University of Engineering & Technology

ACKNOWLEDGEMENT

At first I would like to pray to Almighty Allah to give me the scope & enthusiasm for successful completion of my thesis work.

Firstly I want to express my sincere appreciation and deepest sense of gratitude respect to my supervisor **Professor Dr. Md. Robiul Islam**, Head of Computer Science & Engineering Department, Rajshahi University of Engineering & Technology, Rajshahi, who not only supervised me but also his guidance, advice, continuous encouraging, kind help & the thoughts he has offered have enriched my thesis without which, this thesis work would not have been materialized in the present form.

I am thankful to our respective teachers of the Department of Computer Science & Engineering, Rajshahi University of Engineering & Technology, Rajshahi for their valuable suggestions, extending facilitation & inspiration from time to time.

I am cordially thankful to laboratory staff of CSE department for their co-operation & amiable behavior.

Finally, a lot of thank to all my well-wisher.

November, 2018 Md. Rezaul Karim
RUET, Rajshahi

Heaven's Light is Our Guide

DEPARTMENT OF COMPUTER SCIENCE & ENGINEERING

Rajshahi University of Engineering & Technology, Bangladesh

CERTIFICATE

*This is to certify that this thesis report entitled **"Face Recognition using Viola-Jones Face Detector & Principal Component Analysis Approach"** submitted by **Md. Rezaul Karim, Roll: 133097** in partial fulfillment of the requirement for the award of the degree of Bachelor of Science in Computer Science & Engineering of Rajshahi University of Engineering & Technology, Bangladesh is a record of the candidate own work carried out by him under my supervision. This thesis has not been submitted for the award of any other degree.*

Supervisor	External Examiner
--	--
Professor Dr. Md. Rabiul Islam Head Department of Computer Science &Engineering Rajshahi University of Engineering &Technology Rajshahi-6204	**Firoz Mahmud** Assistant Professor Department of Computer Science &Engineering Rajshahi University of Engineering &Technology Rajshahi-6204

ABSTRACT

Now-a-days face recognition system is widely used in various purposes. A face recognition system is a technology that is capable of identifying or verifying a person from a digital image or video. The working principle of most of the face recognition system is to compare selected facial features from a given image with faces within a database. So, facial portion of a human image is the point of interest. A real-time face recognition system takes a human image as an input. Then the system detects the facial portion and finally recognition procedure is applied. In this thesis work, a system is proposed which consists of both detection & recognition of face. Detection process is performed by Viola Jones face detection algorithm & recognition phase is performed by Principal Component Analysis (PCA). The result of the face detection phase is the facial portion. The features of the face images are extracted & reduced by Principal Component Analysis approach (PCA). This algorithm treats face recognition as a two dimensional recognition problem, taking advantage of the face that faces are normally upright and thus may be described by a small set of 2-D characteristics views. Thus, PCA achieves the feature vectors. These feature vector of the train face images have been trained by K-nearest Neighbor classifier. The proposed face recognition system has been performed on a face Dataset. The PCA approach has also been applied on ORL face dataset. The result highlights the fact the faces need to be frontal & centrally aligned to achieve a good result by Principal Component Analysis Approach.

CONTENTS

LIST OF TABLES

LIST OF FIGURES

Chapter 1
Introduction

1.1 Introduction to Biometrics

Biometrics is a way of establishing the uniqueness of an individual based on the physical, chemical or behavioral properties of the person. This approach is used to identify people based on their biological traits. It is the computation & matching of biological characteristics such as fingerprint images, hand geometry, facial recognition etc. So, biometric system uses different types of physical or behavioral characteristics such as face, retina, fingerprint, voice, face etc.

1.2 Biometrics Consideration

Any human characteristic that satisfies the following conditions can be considered as biometric trait in an application [1].

- **Universality:** Every individual accessing the application possess the trait.
- **Uniqueness**: The given trait should be sufficiently different across individuals.
- **Permanence**: The Biometric trait should be invariant over a period of time.
- **Measurability**: It should be possible to acquire & digitize the biometric trait using suitable devices that do not cause undue inconvenience to the individual.
- **Performance**: The recognition accuracy & the resources applied to achieve that accuracy should meet the constraints imposed by the application.
- **Acceptability**: Individuals in the target population that will utilize the application should be willing to present their biometric trait to the system.
- **Circumvention**: This reflects the security of the biometric system.

1.3 Biometric Modalities

All biometrics may not satisfy the requirement of all identification or authentication problems. Each biometric based authentication system has its strengths and limitations. The following modalities are commonly employed for biometric system.

1.3.1 Face

This modality is one of the most widely used by people to recognize each other. During its estimation, the human brain has progressed highly specialized areas dedicated to the study of facial images. Indeed, in addition to the once vital necessity of being able to instantly determine whether the unexpected encountered individual is a friend or foe, social interactions also strongly depend on the capability to interpret many subtle facial expressions conveying a great deal of information.

1.3.2 Fingerprint

A fingerprint usually appears as a sequence of dark lines that present the high, peaking portion of the fraction ridge skin, while the valleys between these ridges appear as white space and are the low, shallow portion of the friction ridge skin. Fingerprint identification is based primarily on the minutiae, or the location & direction of the ridge endings & bifurcations along a ridge path. Minutiae points are the local ridge characteristics that occur either at a ridge ending or a ridge bifurcation.

1.3.3 Eye (Iris\Retina)

The iris is the colorful part of the eye between the white (sclera) and the pupil. The properties of the iris that enhance its suitability for use in automatic identification include protection from the external environment, impossibility of surgical modification without risking the vision, and physiological response to light which provides a natural test. Iris recognition is fast & noninvasive.

1.3.4 Voice

Voice is a combination of physical and behavioral biometrics. The features of an individual's voice are based on the shape and size of the appendages (e.g., vocal trusts, mouth, nasal cavities, and lips) that are used in the synthesis of the sound. Automated biometric systems based on the modality are called automatic speaker verification (ASV) or automatic speaker identification (ASI) systems.

1.3.5 Other biometrics

There exists other bioinformatics such as: hand geometry, signature, key stroke dynamics, lip recognition, smile recognition, DNA analysis etc.

1.4 Overview of Biometric System

The basic block diagram of a biometric system is shown in figure 1.1. It comprises a sensor module, feature extraction module, a matching module, and a database module [1].

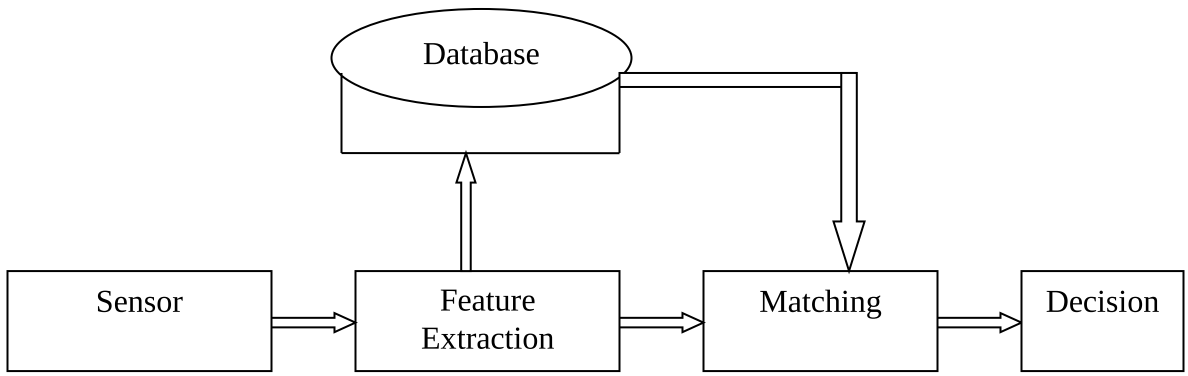


Figure 1.1 Block diagram of a biometric system

1.4.1 Sensor Module

A suitable biometric reader scanner is required to acquire the raw biometric data of an individual. To obtain fingerprint images, for example, an optical fingerprint sensor may be used to take an image of the friction ridge structure of the fingertip.

1.4.2 Feature Extraction module

It is the stage in which the required data is processed to extract feature values. For example, the position & orientation of minutiae points in the fingerprint image would be computed in the feature extraction module of a fingerprint system.

1.4.3 Database module

The database acts as the repository of bioinformatics information. During the enrollment process, the feature set extracted from the raw bioinformatics sample (i.e., the template) is stored in the database along with some biographic information (such as name, Personal Identification Number (PIN), address etc.) characterizing the user.

1.4.4 Matching Module

Matching is carried out when the feature values are compared with those in the template by generating a matching score. For example, in this module, the number of matching minutiae between the query & the template can be computed and processed as a matching score.

1.4.5 Decision making module

In decision making phase, the users claimed identity is either accepted or rejected based on the matching score generated in the matching module.

1.5 Face Recognition System Structure

The necessity for personal security & access control is increasing very rapidly. Biometrics is the technology which is expected to replace traditional authentication methods that are easily stolen, forgotten and duplicated. Fingerprints, face, iris, voice are commonly used biometric features. Among these features, face provides a more direct & convenient identification method & is more acceptable compared with the other biometric features. Although face recognition is not as accurate as other recognition method such as fingerprints, it still grabs huge attention of many researchers in the field of computer vision. The main reason behind this attention is the fact that the face is the conventional way people use to identify each other.

The input of a face recognition system is always an image or video. The output is an identification or verification of the subject or subjects that appear in the input. The face recognition system can be divided into three steps, such as: face detection, feature extraction & classification. Face detection is defined as the process of extracting faces from images. Feature extraction involves obtaining relevant facial features from data. These features could be certain face regions, variations etc. Based on these feature, classification is performed to predict the label of the input image.

Figure 1.2: A generic face recognition system.

1.5.1 Human Image

This is the input of face recognition system. The acquisition of human image can be performed by digital scanning of an existing photograph or by using an electro-optical camera to acquire a live picture of a subject.

1.5.2 Face Detection

Nowadays some applications of face recognition do not require face detection. In some cases, face images stored in databases are already normalized. There is a standard image input format, so there is no need for a detection step. An example of this could be a attendance system database.

However, the conventional input image of computer vision systems is not that suitable. They can contain many items or faces. In these cases face detection is mandatory. Face detection must deal with several well-known challenges. They are usually present in images captured in uncontrolled environments, such as surveillance video systems. These challenges are described below.

- Pose Variation: Pose variation can happen due to subject's movements or camera's angle.
- Feature occlusion: Faces can be partially covered by other objects. The presence of glasses, hats, beards introduces high variability.
- Facial Expression: Different types of facial gestures can cause a variation in facial features.
- Imaging Conditions: Different cameras & ambient conditions can affect the quality of an image.

Viola-Jones face detection can satisfy most of the above criteria for face detection [2]. But it has some limitations also.

1.5.3 Feature Extraction

Feature extraction is an important step in face recognition. The better features are extracted, the higher recognition rate is. Feature extraction is to analyze and transform input information contained in a pattern, and extract information parameters, which are unsusceptible to random interference as the pattern features. The intension is also to eliminate redundant information, improve recognition precision, reduce calculation quantity, and increase calculation speed. Proper features should possess separability, stability & independence. Separability refers to the fact that there exist differences among features of different class. The greater the differences are better for classification. Stability means the features of different pattern of same class should be close. The closer, the better is. In addition, the features are less influenced by random interference. Independence refers to the fact that each feature chosen in one pattern is irrelevant to another. Image features can be classified into following categories

- Visual features.
- Statistical features.
- Transformation coefficient features.
- Algebraic features.

1.5.3.1 Feature Extraction Methods

There are many feature extraction methods**.** Many algorithms have been used, modified & adapted. Some mostly used algorithms are described below.

- **Principal Component Analysis (PCA)** is based on eigenvectors, linear map etc. PCA face recognition does not consider the eye, nose etc. It transforms face images into a set of 'Eigen faces' where most variations exist.
- **Kernel PCA** is based on eigenvectors, non-linear map, kernels.
- **Linear Discriminant Analysis (LDA)** is based on eigenvectors, supervised linear map etc.
- **Independent Component Analysis (ICA)** is a linear map & non-Gaussian based approach.
- **Neural Network** based methods which diverse neural networks using PCA.

- **Gabor Wavelet Transforms** are based on biologically, motivated, linear filter etc.

1.5.4 Classification

Once the features are extracted & selected, the next step is to classify the image. Appearance-based face recognition algorithms use a wide variety of classification methods. Sometimes two or more classifiers are combined to achieve better results. On the other hand, most label-based algorithms match the samples with the model or template. Then, a learning algorithm can be used to improve the algorithm.

Classification methods are used in many areas like data mining, finance, signal decoding, voice recognition, natural language processing or medicine.

Classification algorithms usually involve some learning – supervised, unsupervised or semi-supervised. Consequently, most face recognition systems implement supervised learning methods. The key in building a classifier is similarity, probability & decision boundaries.

- Similarity: This approach is intuitive & simple. Patterns that are similar should belong to the same class. For example, the metric can be the Euclidean distance. Its classification performance is usually good. Some similarity-based classifiers are Template matching, Nearest Mean, KNN etc.
- Probability: Some classifiers are built on a probabilistic approach. Bayes decision rule is often used. Some classifiers with a probabilistic approach are Bayesian, Logistic classifier etc.
- Decision Boundaries: The main idea behind this approach is to minimize a criterion between the candidate pattern & the testing pattern. There are well-known decision boundary based methods, including Fisher Linear Discriminant (FLD), Binary Decision Tree, Perception, Multi-layer Perceptron, Support vector machine etc.

1.6 Applications of Face Recognition

Face recognition has become an important issue in many applications such as security system, credit card verification, in the area of surveillance, closed circuit television &

criminal identification. This system is also useful in human computer interaction, multimedia, online banking, driver licenses, automatic identity verification, home video surveillance system, investigation, personal security, passports, medical records, database recovery, border controls, driver monitoring system, virtual reality etc.

1.6.1 Face Identification

Face recognition systems identify people by their face images. Face recognition systems establish the presence of an authorized person rather than just checking whether a valid identification (ID) or key is being used or whether the user knows the secret personal identification numbers (pins) or passwords. The following are example. To eliminate duplicates in a nationwide voter registration system because there are cases where the same person was assigned more than one identification number. The face recognition system directly compares the face images of the voters & does not use ID numbers to differentiate one from the others. When the top two matched faces are highly similar to the query face image, manual review is required to make sure they are indeed different persons so as to eliminate duplicates [3].

1.6.2 Access Control

In many of the access control applications, such as office access or computer logon, the size of the group of the people that need to be recognized is relatively small. The face pictures are also caught under natural conditions, such as frontal faces and indoor illumination. The face recognition system of this application can achieve high accuracy without much co-operation from user. The following are the example. Face recognition technology is used to monitor continuously who is in front of a computer terminal. It allows the user to leave the terminal without closing files & logging out. When the user leaves for a predetermined time, a screen saver covers up the work & disables the mouse & keyboard. When the user comes back & is recognized, the screen saver clears & the previous session appears as it was left. Any other user who tries to logon without authorization is denied [3].

1.6.3 Security

Today more than ever, security is a primary concern at airports & for airline staff office & passengers. Airport protection systems that use face recognition technology have been implemented at many airports around the world. The following are the two examples. In October, 2001, Fresno Yosemite International (FYI) airport in California deployed Viisages face recognition technology for airport security purposes. The system is designed to alert FYI's airport public safety officers whenever an individual matching the appearance of a known terrorist suspect enters the airport's security checkpoint. Anyone recognized by the system would have further investigative processes by public safety officers. Computer security has also seen the application of face recognition technology. To prevent someone else from changing files or transacting with others when the authorized individual leaves the computer terminal for a short time, users are continuously authenticated, checking that individual in front of the computer screen or at a user is the same authorized person who logged in [3].

1.6.4 Image Database Investigations

Searching image database of licensed drivers benefit receipts, missing children, immigrants & police bookings.

1.6.5 General Identity Verification

There are some application area i.e. Electoral registration, banking, electronic commerce, Identifying newborns, national IDS, passports, employee IDs where human faces used as general identity verification [3].

1.6.6 Surveillance

Like security applications in public places, surveillance by face recognition systems has a lower user satisfaction level, if not lower. Free lighting conditions, face orientations & other divisors all make the deployment of face recognition systems for large surveillance a challenging task. The following are some example of face-based surveillance. To enhance town center surveillance in Newham borough of London, this has 300 cameras linked to the closed circuit TV (CCTV) controller room. The city council claims that the technology has helped to achieve a 34% drop in crime since its facility. Similar systems are in place in Birmingham, England. In 1999 Visionics was awarded a contract from National Institute of Justice to develop smart CCTV technology [3].

1.7 Goal of this Thesis Work

PCA face recognition works by reducing the dimensions of the face space. By applying PCA we can achieve those components where most variations exist among the component. But In this thesis, rather than taking the faces, we are considering the total picture of human face. So face detection also needed. The main goals of this thesis work are listed below:

- To understand how Viola Jones Face Detection works.
- To detect a face by Viola Jones Face Detection.
- To prepare the detected face for recognition.
- To apply PCA to the detected face.
- To select main components based on cumulative co-variance.
- To classify the selected components by KNN.
- To analysis the performance for different databases.

1.8 Organization of this Thesis Work

The thesis work is organized into different chapters to provide an efficient way to understand this work.

In chapter 1, definitions, working principles of biometric system & face recognition system are discussed. An overview of the applications of the face recognition is also given. The goal of this thesis work is also described right at the end of this chapter.

In chapter 2, different types of face detection & recognition system have been included. Block diagram of those recognition systems have also been included.

In chapter 3, Background study related to this thesis work has been described.

In chapter 4, The proposed methodology of this thesis has been elaborated. The block diagram, algorithm, flowchart etc. are also included in this chapter.

In chapter 5, The implementation of the face recognition & detection process has been shown. Based on different criteria, the different implementations are also analyzed.

In chapter 6, The results & performance analysis of this thesis work has been included. The different results are shown tabular form.

In chapter 7, In the conclusion chapter the advantages & disadvantages of the proposed methodology has been described. The limitations & future works have also been discussed.

1.9 Conclusion

In this chapter the introductory topics of a face recognition system have been discussed. Humans have tremendous ability to recognize faces. They can recognize faces even considering aging factors, illumination changes, rotated faces, pose variations, occlusion, expression variations etc. But, for computer to build a face recognition system is quite a challenging task. Because these factors described earlier cause a variations in the normal representation. So to build a face recognition system several steps have to be taken very carefully. The first step is to represent faces in an effective way. Next, is to extract features from those face images. The third & final step is to classify those images. In the following chapters we discussed about those procedures.

Chapter 2
Literature Review

2.1 Introduction

The researchers of computer vision have been attracted by face detection & recognition more & more attention in recent years. With the hardware & software improved the accuracy of the detection is getting better & better. However, the accuracy still can not meet all the requirements. There are many influencing factors when we design a recognition system such as brightness, background, occlusion etc. The face recognition is not a simple process. It is easy for humans to find by their eyes but not easy for a computer to do so. Although very reliable methods of biometric personal identification exist for example fingerprint matching, retina or iris scan, face recognition is very popular among the researchers. Because all the biometric methods involve user participation whereas a personal identification system based on the analysis of profile images of the face does not require user participation directly. Face detection is also the step stone to all the facial analysis algorithms, including face alignment, face modeling, face relighting, face recognition, face verification/authentication, head pose tracking, facial expression tracking/recognition, gender/age recognition, and many more. Researchers have shown interest in the algorithms described above. Some of them are described in the following section.

2.2 Literature Review

Early efforts in face detection have dated back as early as the beginning of the 1970s, where simple heuristic and anthropometric techniques were used [4]. These techniques are largely rigid due to various assumptions such as plain background, frontal face-a typical passport photograph scenario. To these systems, any change of image conditions would mean fine-tuning, if not a complete redesign. Despite these problems the growth of research interest remained stagnant until the 1990s [5], when practical face recognition and video coding systems started to become reality.

Automatic face detection is the cornerstone of all applications which involve facial image analysis including, but no limit to, face recognition and verification [5], face tracking for surveillance [6], facial behavior analysis [7], facial attribute recognition [8] (i.e., gender/age recognition [9]. Automatic face detection was one of the first computer vision applications before this century. Unfortunately the majority of the early works could not performance in unconstrained conditions in real world applications. But this was the first achieve by the work of viola and jones on boosting face detection algorithm [10]. Face detection methods can grouped into five categories: Skin color model-base, Feature invariant, Knowledge-based, Template-based, Edge based.

Viola Jones face recognition algorithm made a great impact in 2000's. It was most likely be the seminal work by Viola and Jones. The viola jones face detector contains three major components which make it possible to build a successful face detector that can detect the face in real world application: the integral image, Adaboost classifier, cascade structure. The Integral image, also known as a summed area table, is an algorithm for quickly and efficiently computing the sum of values in a rectangle sub-set of a grid. The Integral image at any pixel location (x, y) is the sum of the pixels above and to the left and inclusion. The Integral image allows the viola jones face detector to compute the simple Haar-like rectangular features very quickly which is used. The second one is a classifier which has been developed based on Adaboost learning algorithm to select a small number of necessary features from a very large set of features. The third component is cascading which is used to discard background sub images in early processing stages which make the computation more first and increased performance while radically reducing computation time.

In [10] the cascade classifier was trained to detect frontal upright faces. To train the detector, a set of face and non-face training images were used. Viola jones face detector fails significantly on occluded faces. If the eyes are occluded for example, detector will usually fail.

The Adaboost (Adaptive Boosting) algorithm is generally considered as the first step towards other practical boosting algorithms [11]. Viola jones face detection algorithm used Adaboost algorithm as a classifier. In this section, following [12], briefly represent a generalized version of Adaboost algorithm, usually referred as RealBoost. It has been proved by advocated in various works in [13, 14] that RealBoost provides better performance than original Adaboost algorithm.

The cascade structure is a critical component of a viola jones face detector. The key insight is that smaller, and thus more efficient, boosted classifier can be built which rejects most of the negative sub windows while keeping almost all the positive subwindows. As a result majority of the sub windows rejected in the early stages of the detector, making the decision process extremely efficient. The overall process of classifying sub windows thus forms a generate decision tree which is called “cascade” [10]. Thus input sub windows have to pass a series of node during the processing of detection. Each mode makes a binary decision according to which the window will be keep for next round or reject immediately. The numbers of weak classifiers usually increases as the number of nodes a sub-window passes. In [10], the first five nodes contain 1, 10, 25, 25, 50 weak classifier. Each nodes tries to reject a certain amount of negative sub windows while keeping all the positive windows. Having fewer weak classifier at the early stages of face detection also improves the speed of the face detector. The cascade structure also has an effect on the training process. Once argument behind such a process to force the addition of some nonlinearity in the training process which could have improve the overall performance.

In [10], the cascade is constructed manually. That is, the number of weak classifiers and the decision threshold for early rejection at each node are both manually specified. This a non-trivial task. If the decision thresholds were set too aggressively, the final detector will be very fast, but the overall detection rate may be affected. On the other hand, if the decision thresholds are set very conservatively, most sub-windows will need to pass through many nodes, making the detector very slow. Combined with the limited computational resources available in the early 2000’s, it is no wonder that training a good face detector can take months of fine tuning.

Rapid expansion is storage and computation resources, appearance based methods have dominated the recent advances in face detection. Such kind of procedure requires a large set of face and non-face example, and adopt certain machine learning algorithms to learn a face model to perform classification. There are two main key issues in this case: what feature should be extracted and which learning algorithm should be applied. The Haar like rectangular features are very efficient to compute because of using integral image. This kind of technology provides satisfactory results for building frontal face detection. It can't detect the face properly when face is roasted. Researchers extend the straight forward features with more variations in the ways rectangle features are combined. The features set which have been used in [9] have been generalized in by introducing 45 degree rotated rectangular features and center surrounded features.

Few researchers noted the limitations of Haar-like feature set [10] for multi-view face detection and proposed some extended features set by allowing a more flexible combinations of rectangular regions. In [14], three types of features were defined to detect sub window. In [14], Haar like features were proposed which can capture characteristics of human faces more accurately.

In [14] joint Haar-like features were proposed, which are based on the co-occurrence of multiple Haar-like features. It was claimed that feature co-occurrence can better capture the characteristics of human faces, making it possible to construct a more powerful classifier. The joint Haar-like feature uses a similar feature computation and thresholding scheme, however, only the binary outputs of the Haar-like features are concatenated into an index for 2^F possible combinations, where F is the number of combined features. To find distinctive feature co-occurrences with limited computational complexity, the suboptimal sequential forward selection scheme was used in [15]. The number F also heuristically limited to avoid statistical unreliability.

Joint Haar-like features resembles CART tree which is explored in [16]. CART tree based weak classifier improves the performance across various boosting algorithms but decrease the speed very little. Another variation was proposed in [13] where histogram also used with single Haar like features. In another method was proposed in [17] where a new weak classifier called Bayesian stump was introduced. It is also histogram based classifier, however, the split thresholds of the Bayesian stump are derived from iterative split merge operations instead of being equal distances and fixed. Another limitation of original Haar like feature is it can't detect face under extreme light condition [18]. Another well-known feature set which is robust to illumination variations is called local binary pattern. LBP provides very effective result in the field of face detection [18]. LBP were used under Bayesian and boosting framework for face detection in [19]. In [20] pixel pair was used as a feature. Such pixel based feature computation is faster than Haar like features but the performance of the face detection will be decreased.

Another popular complex feature which is used for face detection is regional such as histogram. Local edge orientation was proposed in [21], which compute the histogram of edge orientations in sub regions of the test windows. The orientation histogram is largely invariant to global illumination changes and it is capable of capturing of geometric properties of face that are difficult to capture with linear edge filters such as Haar like features.

Skin color segmentation technique is another important production for face detection. Skin color plays an important role in face detection system. Different color space: basic color spaces (RGB, Normalize RGB, CIE-XYZ), perceptual color spaces (HIS, HSV, HSL, TSL), orthogonal color spaces (YCbCr, YUV, YES YIQ), perceptually uniform color spaces (CIE-Lab, CIE-Luv) and others (Mixture spaces, color ratio spaces) can be used for face detection.

The RGB color space is the default color space for most available image formats. RGB color model was used for face detection in [22]. The HSV space defines color as Hue- the property of a color that varies in passing from red to green, Saturation- the property of a color that varies in passing from red to pink, Brightness (also called Intensity or Lightness or Value)- the property that varies in passing from black to white. The HSV color model has been used in [23]. YCbCr is one of the most popular color spaces for face detection which has been used in [14, 15]. YCbCr space represents color as luminance (Y) computed as a weighted sum of RGB values, and chrominance (Cb and Cr) computed by subtracting the luminance components from B and R values. A variation of YCbCr color space is also used [15].

So, for detecting the human faces different methods have been used. Some of them are feature based approach & some of them are image based approach. But some of the approaches have some limitations. For example, skin color segmentation based approach can not detect faces when the image is taken in poor lighting conditions. Also the skin color approach can not detect faces when there is too much body exposure. Performance of this approach decreases when image is captured from long distance. On the other hand, Viola-Jones face detector has also some limitations. It works well for frontal faces but performance decreases when rotated images are taken into consideration. After the detection is completed, features are extracted by PCA which gives eigenfaces [24].

A general classification of face detection methods is given below:

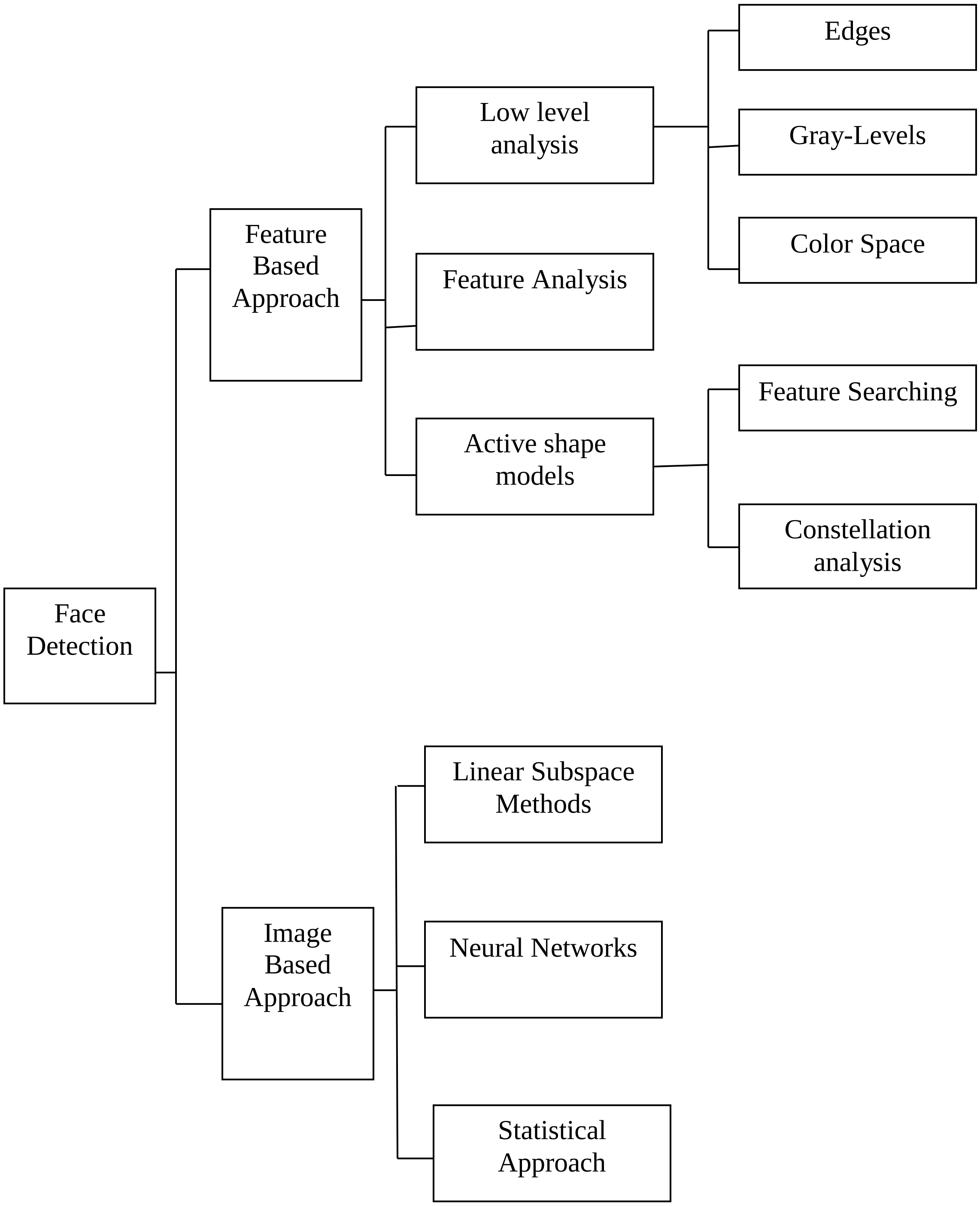


Figure 2.1: Methods of face detection [25]

For face recognition local Binary Pattern of T-zone face area are also taken into consideration before applying PCA [24]. T-zone face area is taken as a matter of interest. The block diagram of T-zone LBP face recognition is given below:

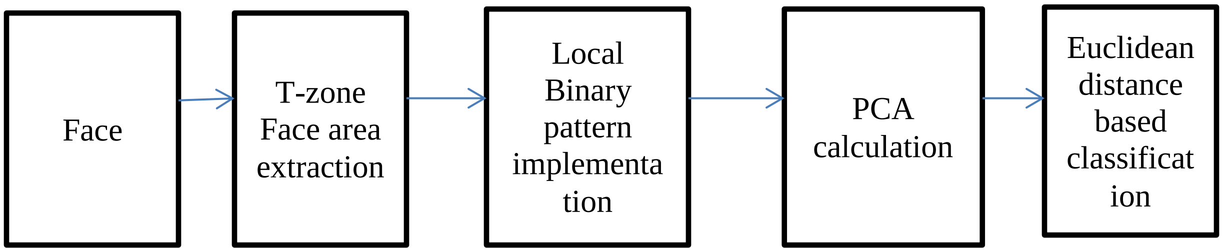


Figure 2.2: Face Recognition by combining LBP & T-Zone [26].

After extracting the principal components several researchers have used probabilistic classifiers. A face recognition model has been developed based on Naive-Bayes classifier [27]. The block diagram of this model is given below:

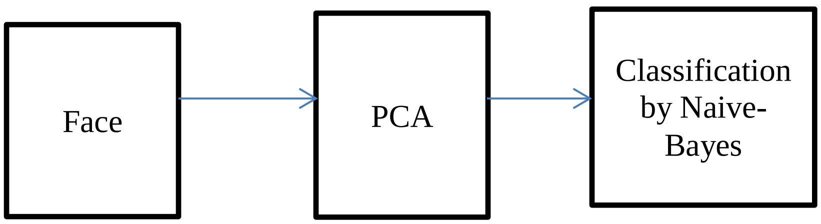


Figure 2.3: Face recognition model using Naive-Bayes Classifier [27].

2.3 Conclusion

In this chapter various face detection methods have been discussed. Some of them are appearance based and others are feature based. The appearance based methods depend on a set of delegate face images to find out face models. The feature based methods try to locate faces by extracting features of the face. There are some face recognition models which also have been discussed in this chapter. They are based on Principal Component Analysis (PCA) Approach.

Chapter 3
Background Study

3.1 Introduction

This thesis work consists of face detection & recognition method. In detection purpose we have used viola-jones face detection algorithm. Features are extracted by Principal Component Analysis. Classification is performed by K Nearest Neighbor approach. To understand the basic operations of these algorithm several image processing, statistical, data mining methods have to be understood. Those basic methods which are related with this thesis work have been described in the following sections.

3.2 Basic concepts of Image representation

We denote images by two dimensional functions of the form f(x, y). The value or amplitude of f at spatial coordinates (x, y) is a positive scalar quantity whose physical meaning is determined by the source of the image. When an image is generated from a physical process, its intensity values are proportional to energy radiated by a physical source (e.g., electromagnetic waves). As a consequence, f(x, y) must be nonzero & finite; that is

$$0 < f(x, y) < \infty$$

The function f(x, y) may be characterized by two components: (1) the amount of source illumination incident on the scene being viewed & (2) the amount of light illumination reflected by the objects in the scene. Appropriately, these are called the illumination & reflectance components and are denoted by i (x, y) & r(x, y) respectively. The two functions contribute as a product to the form f(x, y):

$$f(x, y) = i(x, y) \cdot r(x, y)$$

where $0 < i(x, y) < \infty$ and $0 < r(x, y) < 1$.

Above conditions indicate that reflectance is bound by 0 (total absorption) & 1 (total reflectance). The nature of i (x, y) is determined by the illumination source and r (x, y) is determined by the characteristics of the image objects. These values are presented a matrix from.

The outputs of this function are continuous values. To create a digital image, we need to convert the continuous values into digital form. This involves two processes: sampling & quantization. Digitizing the co-ordinate values is called sampling. Digitizing the amplitude values is called quantization. By the combination of sampling & quantization we achieve a digital image.

3.3 Spatial Correlation & Convolution

There are two closely related concepts that must be understood clearly when performing digital image processing operations. One is correlation & other is convolution. Correlation is the process of moving a filter mask over the image & computing the sum of products at each location. The mechanics of convolution are the same, except that filter is first rotated by 180^o.

3.4 Edge detection

Edge detection is an image processing technique for finding the boundaries of objects within images. It works by detecting discontinuities in brightness. Edge detection is used for image segmentation & data extraction in areas such as image processing, computer vision, machine learning etc.

Generally edges are of three types, such as: horizontal edges, vertical edges, diagonal edges.

Most of the shape information of an object is enclose in edges. At first, when we detect these edges in an image and by using these filters & then enhancing those areas of image which contains edges, sharpness of the image will increase & image will become clear. There are some of the masks for edge detection, such as: Prewitt operator, Sobel Operator, Laplacian operator etc.

3.4.1 Prewitt Operator for edge detection

Prewitt operator is used for edge detection in an image. It detects two types of edges, such as: Horizontal edges & Vertical edges.

Edges are calculated by using difference between corresponding pixel intensities of a image. All the masks that are used for edge detection are also known as derivative masks.

Because as we have stated many times before in this series of tutorials that image is also a signal so changes in a signal can only be calculated using differentiation. So that's why these operators are also called as derivative operators or derivative masks.
All the derivative masks should have the following properties:

- Opposite sign should be present in the mask.
- Sum of mask should be equal to zero.
- More weight means more edge detection.

Prewitt operator provides us two masks one for detecting edges in horizontal direction and another one for detecting edges vertical direction.

-1	0	1
-1	0	1
-1	0	1

Table 3.1: Vertical Edge Detector

Above mask will find the edges in vertical direction and it is because the zeros column in the vertical direction. When you will convolve this mask on an image, it will give you the vertical edges in an image.

When we apply this mask on the image it prominent vertical edges. It simply works like as first order derivate and calculates the difference of pixel intensities in an edge region. As the center column is of zero so it does not include the original values of an image but rather it calculates the difference of right and left pixel values around that edge. This increase the edge intensity and it becomes enhanced comparatively to the original image. The operator which is described above can detect edges in horizontal direction. There is another operator which can also detect edges in vertical direction which is described in the following descriptions.

-1	-1	-1
0	0	0
1	1	1

Table 3.2: Horizontal Edge Detector

Above mask will find edges in horizontal direction and it is because that zeros column is in horizontal direction. When you will convolve this mask onto an image it would prominent horizontal edges in the image.

This mask will prominent the horizontal edges in an image. It also works on the principle of above mask and calculates difference among the pixel intensities of a particular edge. As the center row of mask is consist of zeros so it does not include the original values of edge in the image but rather it calculate the difference of above and below pixel intensities of the particular edge. Thus increasing the sudden change of intensities and making the edge more visible. Both the above masks follow the principle of derivate mask. Both masks have opposite sign in them and both masks sum equals to zero. The third condition will not be applicable in this operator as both the above masks are standardize and we can’t change the value in them. In the following figures the implementation of both the horizontal and vertical edge detectors are illustrated. When we apply vertical mask, all the vertical edges are more visible than the original image and when we have applied the horizontal mask and in result all the horizontal edges are visible.

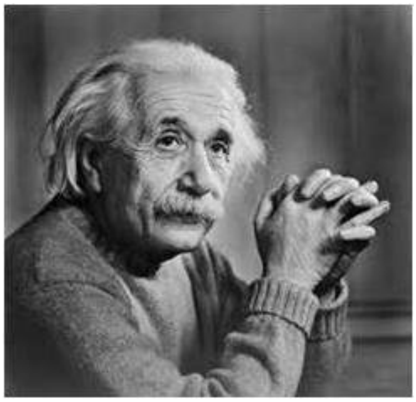


Figure 3.1: Original Image

Figure 3.2: Effect of Vertical Edge Detector

Figure 3.3: Effect of Horizontal Edge Detector

3.5 Basics for K-Nearest Neighbor Classification

The K-nearest-neighbor method was first described in the early 1950's. The method is labor intensive when given large training sets, & did not gain popularity until the 1960's when increased computing power became available. It has since been widely used in the area of pattern recognition. K-nearest neighbor is used as nonparametric method for classification & regression in pattern recognition. Nonparametric regression is a set of techniques for estimating a regression curve without making strong assumptions about the shape of the true regression function. In classification & regression model, k closest training patterns in the feature space are existed [19]. The output depends on whether K-NN is used for classification or regression:

- In K-NN classification, a pattern is classified by a majority vote of it's neighbors, and being assigned to the class most common among it's k ($k \in Z^+$) nearest neighbor.
- In K-NN regression, the output is the property value for the pattern. This value is the average of the values of it's K nearest neighbors.

Nearest-neighbor classifiers are based on training by analogy, that is, by comparing a given test tuple with training tuples that are similar to it. The training tuples are described by n-attributes. Each tuple represents a point in an n-dimensional space. In this way, all of the training tuples are stored in an n-dimensional space. When given an unknown tuple, a KNN classifier searches the pattern space for the K training tuples are closest to the unknown tuple. These k training tuples are the k "nearest neighbors" of the unknown tuple.

Closeness is defined in terms of a distance metric, such as Euclidean distance. The Euclidean distance between two points or tuples, say $X_1 = (x_{11}, x_{12}, x_{13},\ldots, x_{1n})$ & $X_2 = (x_{21}, x_{22}, x_{23},\ldots,x_{2n})$, is

$$\text{dist}(X_1, X_2) = \sqrt{\sum_{i=1}^{n}(x_{1i} - x_{2i})^2}$$

3.6 Conclusion

In this chapter the basic idea behind the representation of digital image, correlation and convolution have been explained. The edge detection which plays a major role in different types of face detection algorithm has been also described. The mathematical explanation of K-Nearest Neighbors (KNN) algorithm is given also in this chapter. The KNN algorithm is a simple, easy to implement supervised machine learning algorithm that can be used to solve classification problem. When applying the KNN classifier, we have to keep focus on selecting the value of K. The value of K should be an odd number. Otherwise the prediction of a specific class will not be possible.

Chapter 4

Proposed Methodology

4.1 Introduction

Human's ability to recognize faces is quite remarkable. They can recognize faces even after many years. Variation in illumination, pose, expression, alignment plays a very little role for humans. But building a face recognition system for computer under different conditions (e.g., illumination, variation of pose, different expression etc.) is quite difficult. In this thesis work a face recognition approach is proposed. Rather than working with the facial image, the whole image of the human is taken into consideration. The facial portion then detected by Viola-Jones Face detection & cropped facial portion is then taken into consideration. The features of the face images are extracted by principal component analysis approach which given Eigen faces. The top Eigen faces are selected based on cumulative covariance. For classification KNN classification technique is used. A block diagram of the proposed methodology is given below. The description of the required algorithm is given in the following sections.

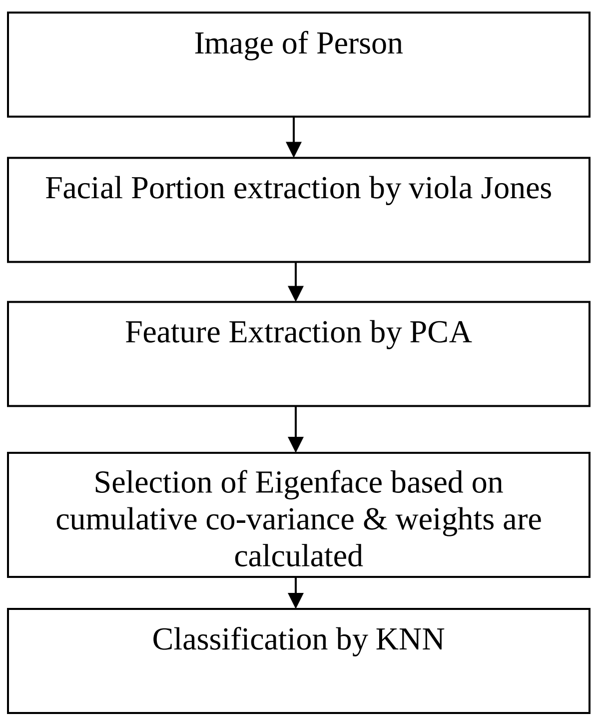


Figure 4.1 Block diagram of proposed Methodology

4.2 Viola-Jones Face Detection

A face detection system is designed by giving input as some faces & non-faces & training a classifier or something that identifies a face. Training is performed using positive & negative faces i.e., faces & non-faces. And once the training is done with the data that we have got we would be able to detect any face in any incoming image for long. We show input images of a face to an alien who has probably no previous knowledge of what a human face is. So, we show several thousands of human faces to tell him/her this is a human face. We also show thousands of non-faces & tell him/her that this is not a face. So basically, the brain tries to identify features by watching all the pictures of human faces & registers & then it also extracts features of non-faces & it registers another subset. So, brain tries to identify what is a face is & what a non-face is using extracted features & once the alien has been trained, then whenever we show a new image to him/her, he/she will be able to identify based on understanding the features of the input image. Here, exactly what we are trying to do to understand what is a face & what is a non-face.

A face detector has to make a decision whether an image of arbitrary size contains a human face and if so, where is the location. The problem of detecting the presence and locations of faces in arbitrary images is a tough one, but it has received significant attention over the last 30 years. The task of detecting a face in an image is not an easy problem because many difficulties arise and must be taken into account. The Viola Jones object detection framework to provide competitive object detection rates in real-time proposed in 2001 by Paul Viola and Michael Jones. Although it can be trained to detect a wide variety of object classes, it was motivated primarily by the problem of face detection. There are three key features in Viola Jones face detection algorithm. The first one is integral image representation which allows the computation of Haar features very quickly. The second one is Adaboost feature selection which selects the best visual features among a set of possible features. The final one is a method for combining classifiers in cascade with a view to giving more importance on promising face like regions.

4.2.1 Haar Features

The features sought by the detection framework universally involve the sums of image pixels within rectangular areas. As such, they bear some resemblance to Haar basic functions, which have been used previously in the realm of image-based object detection. However, since the features used by Viola & Jones all rely on more than one rectangular area, they are generally more complex. Fig 4.2 illustrates the four different types of features used in the framework. The value of any given feature is the sum of feature is the sum of the pixels within shaded rectangles. Rectangular features of this sort are primitive when compared to alternatives such as steerable filters. Although they are sensitive to vertical & horizontal features, their feedback is considerably coarser. All human faces share some similar properties. Eyes, mouth, bridge of nose are the important facial features that needed to be considered. These regularities may be matched using Haar features.

A few properties common to human faces:

- The eye region is darker than the upper checks.
- The nose bridge region is brighter than the eye.

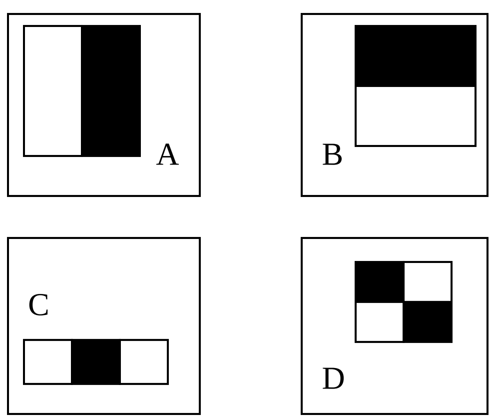


Figure 4.2: Haar Features

Some of the Haar features are edge feature & some are line features. The eyes of a human face contain darker pixels & the forehead right above the eyebrow contains brighter pixels. So, this property of human face is very similar to 4.2.B Haar feature. The nose of a human face contains brighter region than the checks region. So, we can detect the nose region using one of the Haar features which is illustrated in figure 4.3.

Figure 4.3: Eye Detection using Haar Features.

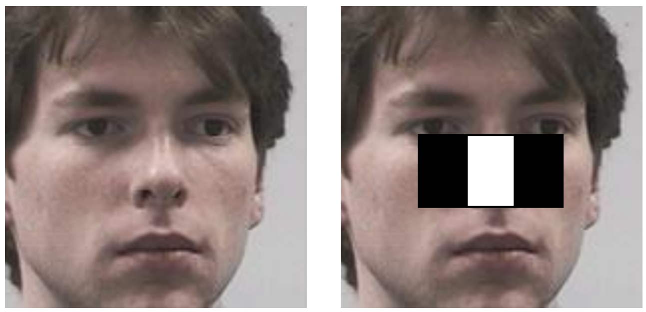


Figure 4.4: Nose Detection using Haar Features.

So, we are able to assign Haar features to the most relevant features on the human face. The Viola Jones Face Detection algorithm is going to detect human faces based on the most relevant features such as: eyes, noses, lips, eyebrows etc. This is why Haar features are extremely important in detecting human faces.

$$\nabla = \frac{1}{n}\Sigma(\text{pixels in the black area}) - \frac{1}{n}\Sigma(\text{pixels in the white area})$$

For ideal case the ∇ is 1. The value will not be 1 in real case scenario. The closer the value two 1, the more likely we have found Haar features [2].

4.2.2 The Integral Image

Every single time we need to sum up all the pixels of the black region & all the pixels of white region to calculate the difference value of Haar features. This does not look very computationally efficient for us when we want to calculate. So Viola Jones algorithm provides a solution or a trick metric to solve the problem that is called an Integral Image.

So, the basic idea of Integral Image is when to need to calculate the sum of a patch, we do not need all the values of that patch, just the corner values of that patch is needed [2].

In an integral image the value at pixel (x, y) is the sum of pixels above & to the left of (x, y).

1	1	1
1	1	1
1	1	1

a. Original

1	2	3
2	4	6
3	6	9

b. After Calculation

Figure 4.5: Integral Image

Integral image allows the calculation of sum of all pixels inside any given rectangle using only four values at the corners of the rectangle.

Sum of pixels in D = 1+ 4- (2 + 3)
= A+ (A+B+C+D) – (A+B+A+C)
= D

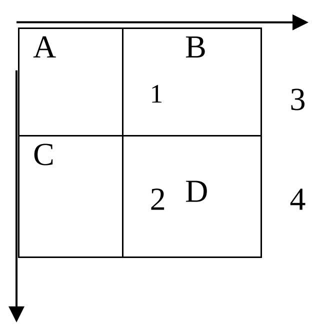


Figure 4.6: Pixel Calculation of Integral Image.

4.2.3 Adaboost Learning

The Haar features have to be evaluated across the entire image, till we end up reaching the last pixel in the image. So, we would have evaluated the features couples of hundreds of times. If we increase the size of the feature & apply this same feature across the image, we have to calculate thousands more. Viola Jones algorithm uses a 24*24 window as the base window to start evaluating these features in a given image. If we consider all possible parameters of Haar features like position, scale, type we will end up calculating approximately 160,000 features in this window [2]. So, right now we have got couple of problems in here. So, to solve the problem the basic idea is to eliminate a lot of features which are redundant or which are not useful & select only the features that are very useful to us. In order to modify these features Viola Jones use a modified version of the Adaboost Algorithm developed By Freund & Schapire in 1996

Adaboost is a machine learning boosting algorithm capable of constructing a strong classifier through a weighted combination of weak classifiers. A weak classifier classifies correctly in only a little bit more than half the cases. To match this terminology to the presented theory each feature is considered to be a potential weak classifier. A weak classifier is mathematically described as

$$H(x, f, p, \theta) = \begin{cases} 1 \text{ if } pf(x) > p\theta \\ 0 \text{ otherwise} \end{cases}$$

Where x is a 24*24 pixel sub window, f is the applied feature, p is the polarity & θ the threshold that decides x should be classified as a positive (a face) or a negative (a non-face).

Since only a small amount of the possible 160,000 feature values are expected to be potentially weak classifiers the Adaboost algorithm is modified to select the best features. Viola Jones' Modified Adaboost Algorithm is presented in the following figure.

An important part of the modified Adaboost algorithm is the determination of the best feature, polarity & threshold. There seems to be no smart solution to this problem & Viola Jones suggest a simple brute force method. This means that the determination of each new weak classifier involves evaluating each feature on all the training examples in order to find the best performing feature. With the integral image, the computationally efficient features & the modified Adaboost Algorithm in place it seems like the face detector is ready for implementation, but Viola Jones have one more ace up the sleeve.

4.2.4 Cascaded classifier

The basic principle of the Viola Jones face detection algorithm is to scan the detector many times through the same image – each time with a new size. Even if an image should contain one or more faces it is obvious that an excessive large amount of the evaluated sub-windows would still be negatives (non-faces). This realization leads to a different formulation of the problem: Instead of finding faces, the algorithm should discard non-faces. The thought behind this statement is that it is faster to discard a non-face than to find a face. With this in mind a detector consisting of only one (strong) classifier suddenly seems inefficient since the evaluation time is constant no matter the input. Hence the need for a cascaded classifier arises. The cascaded classifier is composed of stages each containing a strong classifier. The Job of each stage is to determine whether a given sub-window is definitely not a face or may be a face [2].

When a sub-window is classified to be a non-face by a given stage it is immediately discarded. Conversely a sub-window classified as a maybe-face is passed on to the next stage in the cascade. It follows that the more stages, the higher the chance the sub-window actually contains a face. The concept is illustrated with two stages in the next figure.

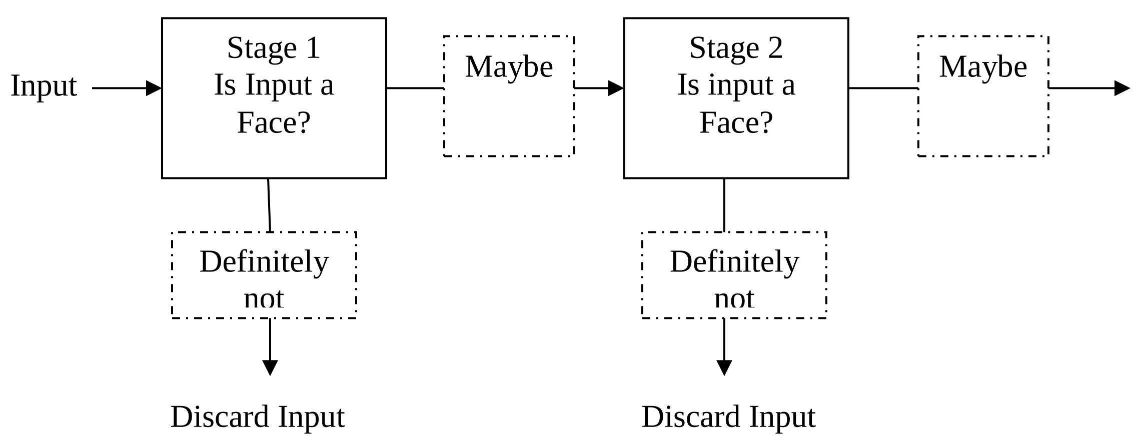


Figure 4.7: Block diagram of Cascaded Classifier

4.3 Principal Component Analysis

Principal Component Analysis (PCA) is a mainstay of modern data analysis. PCA is a kind of dimension reduction or ordination analysis. Ordination analysis attempts to embed objects distributed in high dimensional space into lower dimensional space. In PCA, dimension reduction is achieved by projection to lower dimensional space using linear transformation. Although PCA is a simple and classical method, it can often effectively reduce redundant information. Principal Component Analysis (PCA) has been called one of the most valuable results from applied linear algebra. PCA is used abundantly in all forms of analysis – from neuroscience to computer graphics-because it is simple, non-parametric method of extracting relevant information from confusing datasets. With minimal addition effort PCA provides a roadmap for how to reduce a complex dataset to a lower dimension to reveal the sometimes hidden, simplified dynamics that often underlie it. The important fact which is considered is that although face images have high dimensionality, in reality they span very low dimensional space. So instead of considering whole face space with high dimensionality, it is better to consider only a subspace with lower dimensionality to represent this face space. The 'Eigenface' approach uses Principal Component Analysis (PCA) algorithm for the recognition of the images. It gives us efficient way to find the lower dimensional space. The scheme is based on an information theory approach that decomposes face images into a set of characteristic feature images called 'Eigenface', which are actually the principal components of the initial training set of face images.

4.3.1 An overview of PCA algorithm

First, the original images of the training set are transformed into a set of eigenfaces E. Afterwards; the weights are calculated for each image of the training set & stored in the set W. Upon observing an unknown image X, the weights are calculated for that particular image & stored in the vector WX. Afterwards; WX is compared with the weights of images, of which one knows for certain that they are facing (the weights of the training set W).

One way to do it would be to regard each weight vector as a point in space & calculate an average distance D between the weight vectors from WX & the weight vector of the unknown image WX. If this average distance exceeds some threshold value, then the weight vector of the unknown image WX lies too far apart from the weights of the faces. In this case, the unknown X is considered to not a face. Otherwise (if X is actually a face), its weight vector WX is stored for classification. The optimal threshold value has to be determined empirically. By means of PCA one can transform each original image of the training set into a corresponding Eigen face. An important feature of PCA is that one can reconstruct any original image from the training set by combining the Eigen faces. Remember that Eigen faces are nothing less than characteristic features of faces. Therefore one could say that the original face image can be reconstructed from Eigen faces if one adds up all the Eigen faces (features) in the right proportion. Each Eigen Face represents only certain features of the face, which may or may not be present in the original image. If the feature is present in the original image to a higher degree, the share of the corresponding Eigen Face in the "sum" of the Eigen Faces should be greater.

If, contrary, the particular feature is not (almost not) present in the original image, then the corresponding Eigen Face should contribute a smaller (or not at all) part to be sum of the Eigen Faces. So, in order to reconstruct the original image from the Eigen Faces, one has to build a kind of weighted sum of all Eigen Faces, with each Eigen Face having a certain weight. This weight specifies, to what degree the specific feature (Eigen Face) is present in the original image.

4.3.2 Significance of Eigen Face Approach

Principal Component Analysis (PCA) is a powerful statistical technique that converts a set of correlated variables into a set of linearly uncorrelated variable called principal component. It is a kind of orthogonal transformation. It is a simple, non-parametric method for extracting relevant information from confusing data sets. With minimal effort PCA provides a road map for how to reduce a complex data set to a lower dimension to reveal the sometimes hidden. Simplified structures that often under-lie it. The goal of Principal Component analysis is to identify the most meaningful basis to re-express a dataset. The hope is that this new basis will filter out the noise and reveals hidden structure. Indeed, PCA makes one stringent but powerful assumption: linearity. Linearity vastly simplifies the problem for confine the set of potential bases.

The standard Eigen Face Approach has given a great importance in Face Recognition. Let, M eigenvectors for representing training set images. Now it is important to choose only M` Eigenvectors from these M eigenvectors, such that M` is less than M, to represent face space spanned by images. This will reduce the face space dimensionality & enhance speed for face recognition. Here we are reducing the dimensionality of face images. We can choose only M` Eigenvectors with largest Eigenvalues. Now as the higher Eigenvalues represent maximum face variation in the corresponding Eigenvector direction, it is important to consider this Eigenvector for face image representation. Since, the lower Eigenvectors does not provide much information about face variation in corresponding Eigenvector direction, such small eigenvectors can be neglected to further reduce the dimension of face space. This does not affect success rate much & is acceptable depending on the application of face recognition.

4.3.3 PCA Face Recognition Algorithm Analysis [24]

Step 1: A set of M face images (I_i ; i = 1, 2, …, M) is first required.

$$I_i = \begin{vmatrix} \gamma_{11} & \gamma_{12} & \ldots\ldots\ldots & \gamma_{1N} \\ \gamma_{21} & \gamma_{22} & \ldots\ldots\ldots & \gamma_{2N} \\ \ldots & \ldots & \ldots & \ldots \\ \gamma_{N1} & \gamma_{N2} & \ldots\ldots & \gamma_{NN} \end{vmatrix}_{N*N}$$

for i = 1, 2, 3, ….M. Where [γ_{ij}] $_{i, j= 1, 2, 3, .., N}$ are the gray values of the pixels.

Step 2: Each face image I_i is expressed as a vector Γ_i (i = 1, 2, 3, …, M) by concatenating each row (or column) into a long thin vector as shown in below. Since each image is (N*N) matrix so Γ_i is (N^2*1) vector.

$$\Gamma_i = \begin{vmatrix} \gamma_{11} \\ \gamma_{12} \\ \ldots \\ \gamma_{1N} \\ \gamma_{21} \\ \ldots \\ \gamma_{2N} \\ \ldots \\ \gamma_{N1} \\ \ldots \\ \gamma_{NN} \end{vmatrix}_{N^2 * 1}$$

Step 3: Compute the average face vector of the set, which is given by the following equation.

$$\boldsymbol{\Psi} = \frac{\mathbf{1}}{\mathbf{M}} \sum_{\mathbf{i=1}}^{\mathbf{M}} \boldsymbol{\Gamma}\mathbf{i}$$

$$\Psi = \begin{bmatrix} \psi_{11} \\ \psi_{12} \\ \vdots \\ \psi_{1N} \\ \psi_{21} \\ \vdots \\ \psi_{2N} \\ \psi_{N1} \\ \psi_{N2} \\ \vdots \\ \psi_{NN} \end{bmatrix} = \frac{1}{M} \begin{bmatrix} \gamma_{11}{}^{1} & + & \gamma_{11}{}^{2} & + & \dots & + & \gamma_{11}{}^{M} \\ \gamma_{12}{}^{1} & + & \gamma_{12}{}^{2} & + & \dots & + & \gamma_{12}{}^{M} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \gamma_{1N}{}^{1} & + & \gamma_{1N}{}^{2} & + & \dots & + & \gamma_{1N}{}^{M} \\ \gamma_{21}{}^{1} & + & \gamma_{21}{}^{2} & + & \dots & + & \gamma_{21}{}^{M} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \gamma_{2N}{}^{1} & + & \gamma_{2N}{}^{2} & + & \dots & + & \gamma_{2N}{}^{M} \\ \gamma_{N1}{}^{1} & + & \gamma_{N1}{}^{2} & + & \dots & + & \gamma_{N1}{}^{M} \\ \gamma_{N2}{}^{1} & + & \gamma_{N2}{}^{2} & + & \dots & + & \gamma_{N2}{}^{M} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \gamma_{NN}{}^{1} & + & \gamma_{NN}{}^{2} & + & \dots & + & \gamma_{NN}{}^{M} \end{bmatrix}_{N^2 X1}$$

Where M is the number of training set images.

Step 4: The images are mean centered by subtracting the mean image from each image vector.

$$[\ \mathbf{Ø_i = \Gamma_i - \Psi}\ ]_{N^2 * 1} \qquad \textbf{for i = 1, 2, 3, ….,M.}$$

The purpose of subtracting the mean image from each image vector is to be left with only the distinguishing features from each face & removing in way information that is common.

Step 5: Build the mean-subtract face matrix A.

The mean–subtract faces are arrayed in a matrix A, with one column per sample image.

$$A = \mathbf{[\ Ø_1 \quad Ø_2 \quad Ø_3 \quad …. \quad Ø_N\ ]_{N^2 * M}}$$

Step 6: Getting the eigenvalues & eigenvectors. The covariance matrix is give as:

$$\mathbf{C} = \frac{\mathbf{1}}{\mathbf{M}} \sum_{\mathbf{n=1}}^{\mathbf{M}} (\boldsymbol{\varphi}_\mathbf{n} \boldsymbol{\varphi}_\mathbf{n}^\mathbf{T}) = \mathbf{AA^T}$$

Where C is $N^2 * N^2$.

There covariance matrix C however is $N^2 * N^2$ real symmetric matrix, and determining the N^2 eigenvectors & eigenvalues is an intractable task for typical image sizes. We need a computationally feasible method to these eigenvectors.

Consider the eigenvectors V_i of $A^T A$ such that

$$A^T A V_i = \mu_i V_i$$

Multiplying both sides by A, we have

$$AA^T A V_i = \mu_i A V_i$$

$$C A V_i = \mu_i A V_i$$

$$C U_i = \mu_i V_i$$

From which we see that $A V_i$ are the eigenvectors of $C = AA^T$.

Thus AA^T & A^TA have the same nonzero eigenvalues & their eigenvectors related as follows:

$$U_i = A V_i$$

Following these analysis, we construct the M*M matrix L.

$$L = A^TA$$

Where $L_{mn} = \varphi_m^T \varphi_n$

Step 7: This set of very large vectors is then subjected to principal component analysis, which seeks a set of M orthogonal vectors U_n, which best describes the distribution of the data. The K^{th} vector U_k is chosen such that

$$\boldsymbol{\lambda_k = \frac{1}{M}\sum_{n=1}^{M}(U_k^T \varphi_n)^2}$$

is a maximum, subject to U_l^T & scalars λ_k are the eigenvectors & eigenvalues of C respectively.

Step 8: After sorting eigenvectors based on their eigenvalues, we can analyze that the most significant components. The top-eigenvectors contribute very much. So, based on cumulative co-variance the eigenvectors are chosen for projection purpose.

Step 9: Compute the weights of each face for projecting into the face space.

$$\mathbf{w_{ik}} = \mathbf{U_k^T}(\,\boldsymbol{\Gamma}\mathbf{i} - \boldsymbol{\psi}) \quad \text{for i, k = 1, 2, 3,…,M.}$$

$$\mathbf{w_{ik}} = \mathbf{U_k^T}\boldsymbol{\varphi}_\mathbf{i}$$

where U_k^T are the eigenvectors.

These weights form the feature vectors.

$$\boldsymbol{\Omega}_\mathbf{i}^\mathbf{T} = [\mathbf{w_{i1}} \quad \mathbf{w_{i2}} \quad \mathbf{....} \quad \mathbf{w_{in}}\,] \quad \text{for I = 1 , 2, 3,….,M}$$

The feature vectors describe the contribution of each eigenfaces in representing the input face image, treating the eigenfaces as a basis set of face images.

4.4 Classification By K-Nearest neighbor

The block diagram of face recognition system using K nearest neighbor is shown in the following figure.

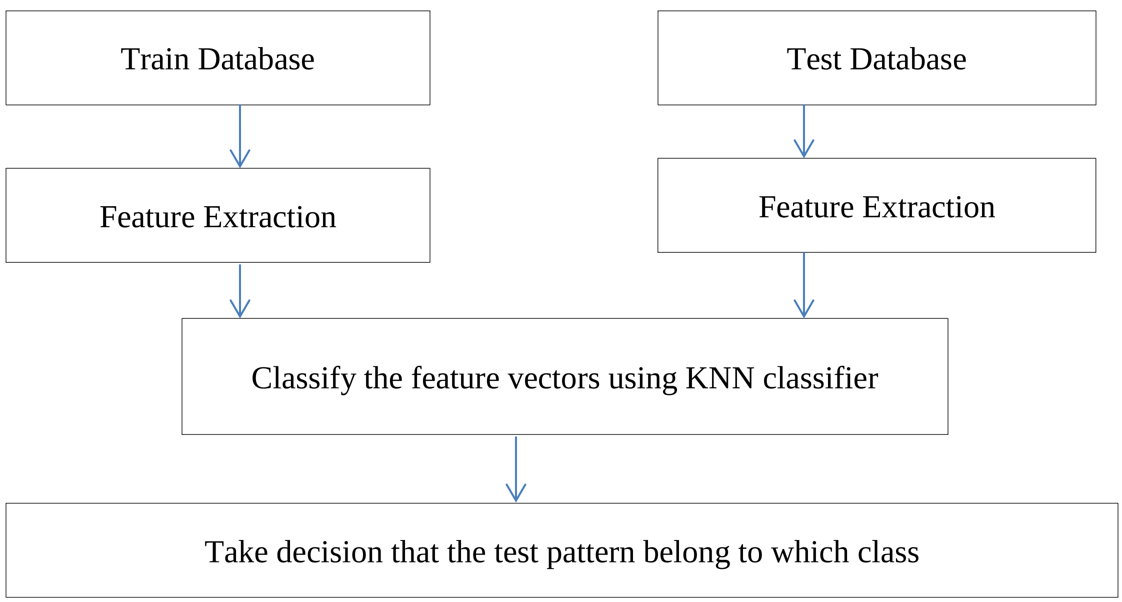


Fig 4.8: Block diagram of face recognition using KNN.

The steps performed to recognize an unknown face based on the learned knowledge using KNN classifier are given below:

1. Feature extraction of train images: Features are extracted from train facial database using PCA technique.
2. Knowledge gather: Using KNN, (k is determined based on face images per person) the features are classified.

3. Feature extraction of unknown face pattern: When an unknown pattern is required to be classified based on gathered knowledge, first the features vector has been extracted according to the technique applied in the training phase.
4. Take decision: The final step is to classify the test face features using KNN & take decision which class the unknown probe belong to.

4.5 Conclusion

In this chapter, the block diagram of proposed methodology is presented. The description of every method is also presented with the mathematical equations. The basic principle of the Viola Jones face detection algorithm is to scan to scan a sub-window capable of detecting faces across a given input image. After the face part is cropped, Principal Component Analysis (PCA) approach has been used for feature extraction. PCA approach treats face recognition as a two dimensional recognition problem, taking advantage of the face that faces are normally upright & thus may be described by a 2-D characteristic views. Face images are projected onto a feature space that best encodes the variation among known face images. The face space defined by the Eigen faces which are the Eigen vectors of the set of faces. After that feature extraction by PCA, we achieve those dimensions where most variations exist. Finally, KNN classifier is applied to predict a test image class.

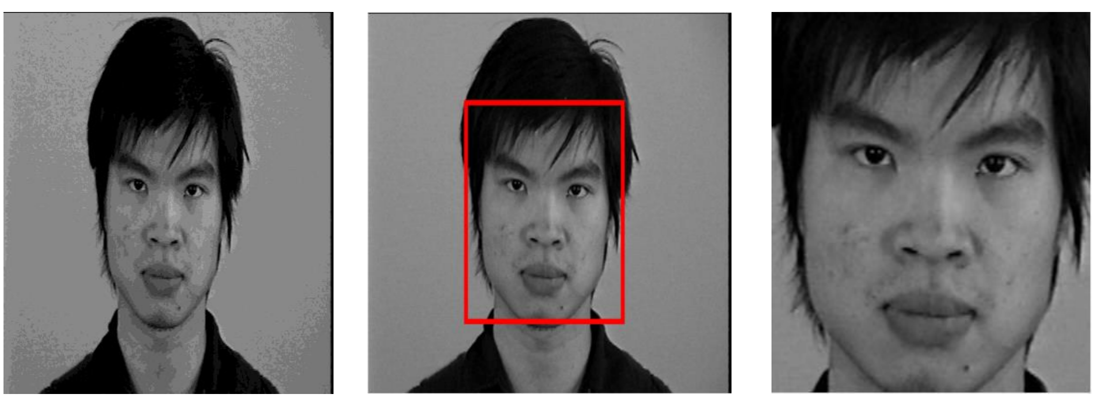
Chapter 5
Implementation

5.1 Introduction

In this chapter we discuss the implementation of proposed methodology. The proposed methodology has been performed on a database [28] which contains human images. The PCA face recognition is applied to ORL [29] face database which contains human faces. So, in ORL face database detection algorithm is not required. The Viola Jones face detector is an efficient face detection algorithm. But there are also some limitations of it. Some of the limitations are inability to work with rotated faces, sensitive to lighting conditions etc. The PCA feature extraction works by reducing the dimension size. Based on cumulative co-variance the features are selected. Classification is performed by KNN classifier.

5.2 Face Detection

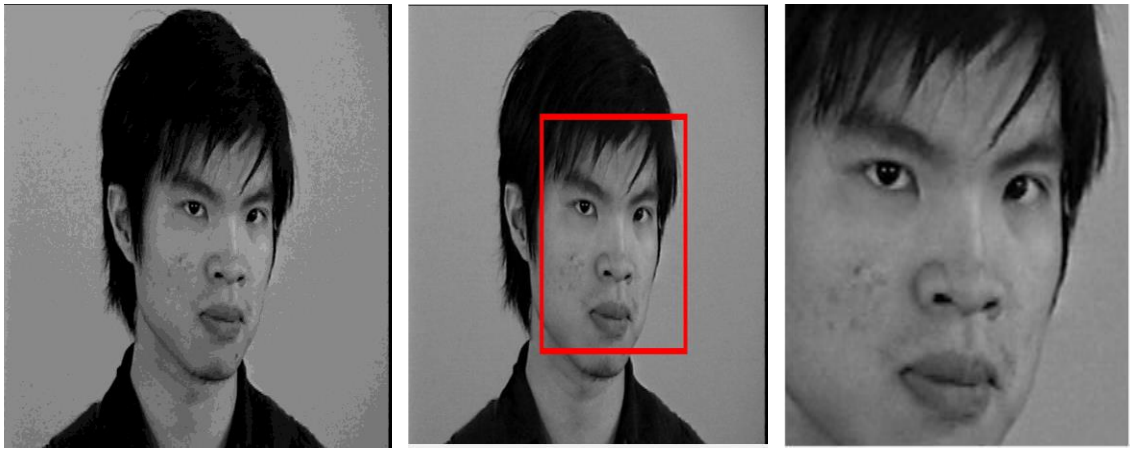
Viola Jones Face Detection is a robust face detection algorithm. For detecting the face from the whole image we have used Haar-like cascaded classifier. Some of the examples of the face detection are given below:



a. Human Image b. Detected face portion c. Cropped portion

Figure 5.1: Face Detection of frontal face by viola jones face Detection Algorithm

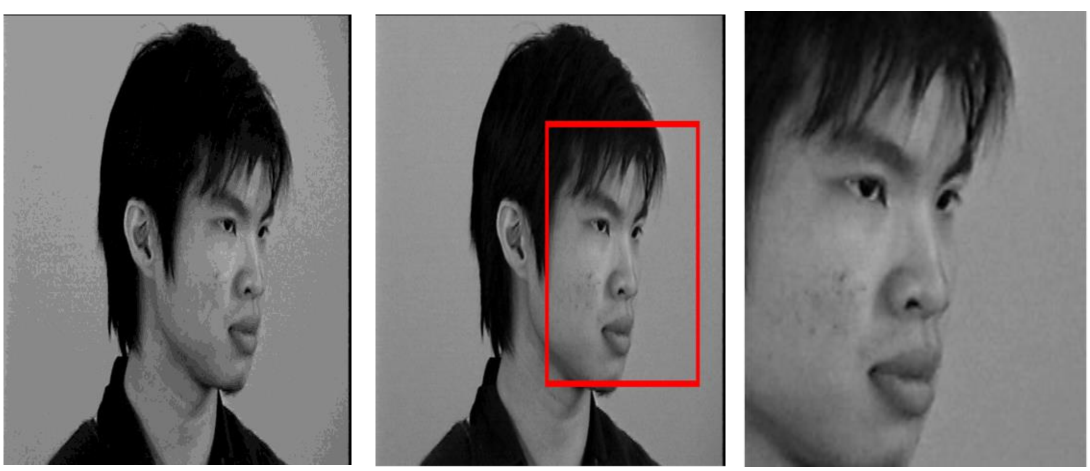
In the figure 5.1 the red box showing the detected face portion and then that portion is cropped. Viola Jones Face detection works well when frontal face are used. If the images are rotated, then the performance of this detection algorithm decreases. Some of the examples of rotated face detection are given in the following figures.



a. Image b. Detected face portion c. Cropped face portion

Figure 5.2: Face Detection of $+20^{o}$ rotated face by viola jones face Detection Algorithm

The viola jones face detection algorithm works almost well for this 20^{o} rotated image. But some additional portions (non-face region) are also included in the cropped version. As the rotation amount increases, the non-face region also increases. This algorithm can not detect the face when the image is 70^{o} rotated. These examples are listed below



a. Image b. Detected face portion c. Cropped face portion

Figure 5.3: Face Detection of $+60^{o}$ rotated face by viola jones face Detection Algorithm

If the face is more rotated, it can not be detected by the viola jones face detection algorithm. In the following figure an image is shown where face is 70^{0} rotated and the Viola Jones face detection algorithm failed to detect that face.

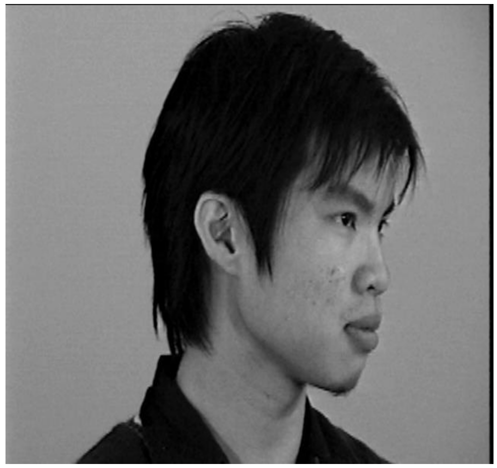


Figure 5.4: $+70^0$ rotated face

So, when faces are rotated the performance of the Viola Jones Face Detection algorithm decreases. It ends either not detecting it or adding some non-face regions.

5.3 Feature extraction

In this thesis work we have used PCA for feature extraction of those faces which are at most 20^o rotated. PCA transforms pattern dataset into machine readable dataset where are different from each other. Similar characteristics of the pattern or face are removed. In this technique the original dataset are projected into new reduced dimension so that the variances of images in the dataset are very high. The steps of PCA calculation are given below:

Step 1: First the mean image of the face images is calculated.

Figure 5.5: Mean Face.

Step 2: Then each image is subtracted from the mean image to remove the common features. The subtracted image can also be called as normalized image. Some of the normalized images are given below.

Figure 5.6: Normalized Faces.

Step 3: The Eigen faces are generated from these normalized images. Eigen faces are sorted based on their eigenvalues.

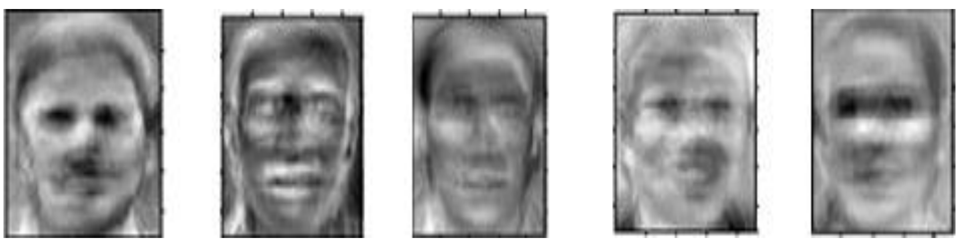


Figure 5.7: Eigen faces.

Step 4: The Eigen faces are selected based on cumulative covariance.

Step 5: The Eigen faces are then multiplied by the normalized images to project them onto the face space.

5.4 Classification by KNN

Step 1**:** The features and then weights of the unknown image are calculated.

Step 2: The Euclidean distance between all the projected face images & the test image is calculated.

Step 3: The shortest k items are selected for consideration.

Step 4: Majority of k items' class is determined as the class of the test image.

5.5 Conclusion

In this chapter, the implementation of proposed methodology is presented. The frontal faces are detected quite effectively by Viola Jones face detection algorithm. But it is quite obvious that when detecting rotated face, the performance decreases. For computing the Eigen faces at first the mean face need to be computed which represents the common properties of the face dataset. Then normalized images of every individual is computed which represents the deviation of every image from the mean image. Finally, the eigenfaces are computed. The weights of every individual is computed for projecting those individuals on to the face space. When a test face image comes, we have to project that faces onto the face space. For class prediction, we have to figure it out that the test image is closer to which individual and that individual's class will be labeled as the test image class.

Chapter 6
Results & Performance Analysis

6.1 Introduction

For performance analysis, we have used two datasets. The first dataset contains 40 different subjects [28]. Each subject of that database contains 70 images of different pose. After the implementation of Viola Jones face detection algorithm, 18 images per subject are taken into consideration for feature extraction. Because faces of other images either not detected by the detector or too many non-face regions were included. The second database that we have worked on is ORL Face dataset [29] where alignment of all the faces is same. We have just applied Principal Component Analysis (PCA) on ORL face dataset.

6.2 Experimental result of Proposed Methodology on Face Dataset:

The input of our proposed methodology is a human image. The face is extracted from that image. Performance of proposed methodology is given below:

Table 6.1: The performance of Proposed Dataset

Cumulative variance	No. of Component	Test Accuracy (%)
<60	9	69.75
<65	12	73.17
<70	16	74.34
<75	37	78.45
<80	46	78.16
<85	77	77.92
<90	157	77.16

Our proposed dataset contains 40 different persons [28]. Each person has 18 images (after detection) and from those 18 images 70% are used for training and 21% are used for testing. For KNN classifier k = 3 is used. There are 40 different people i.e., 40 classes exist here. Maximum recognition rate using KNN classifier is 78.45%, when we select 37 components of training data set.

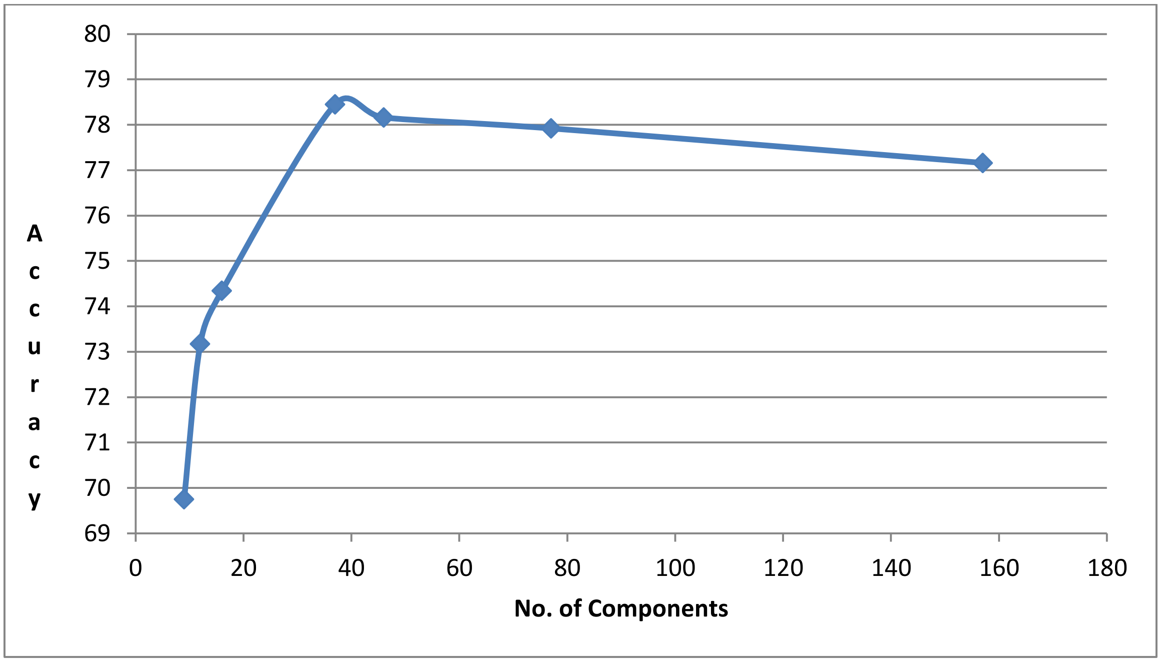


Figure 6.1: The performance of PCA based KNN classifier.

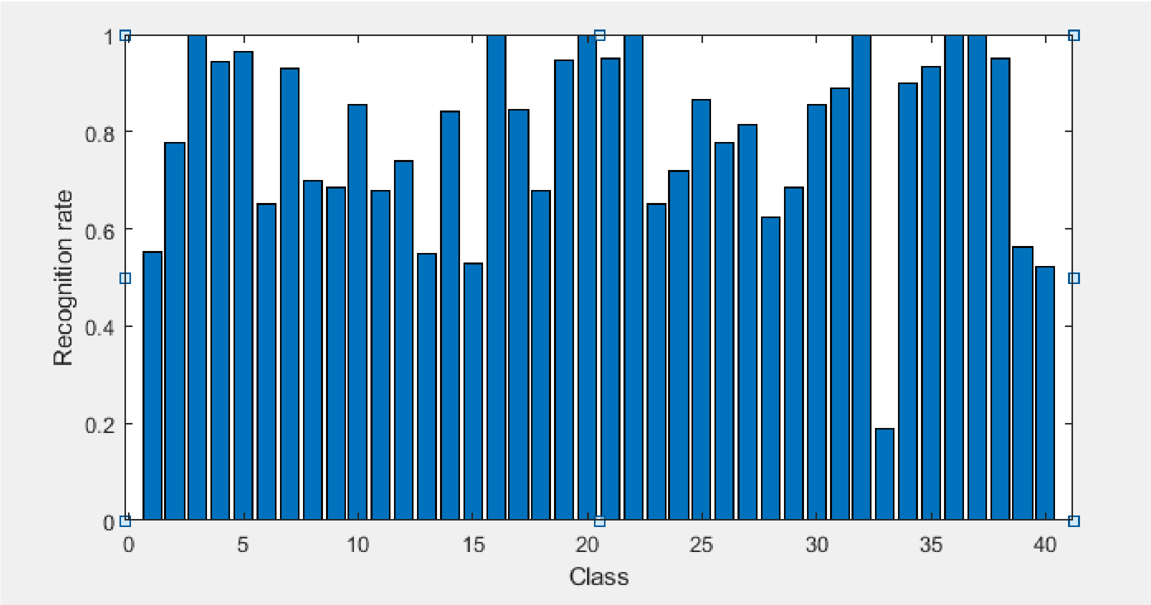


Fig 6.2: Recognition Rate of each class.

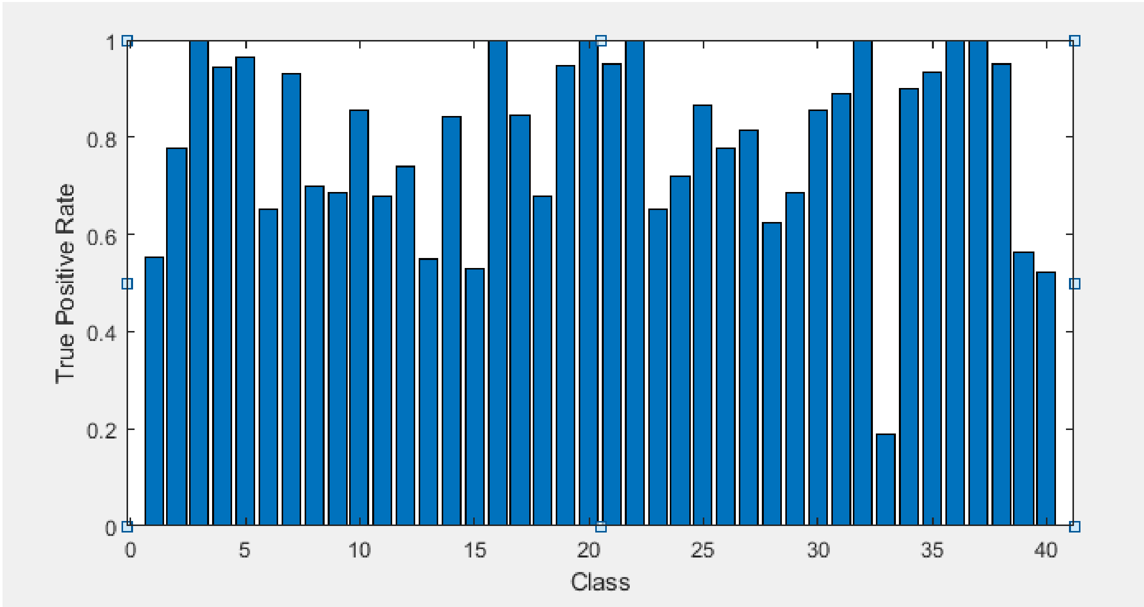


Figure 6.3: Sensitivity Rate for each class of KNN classifier

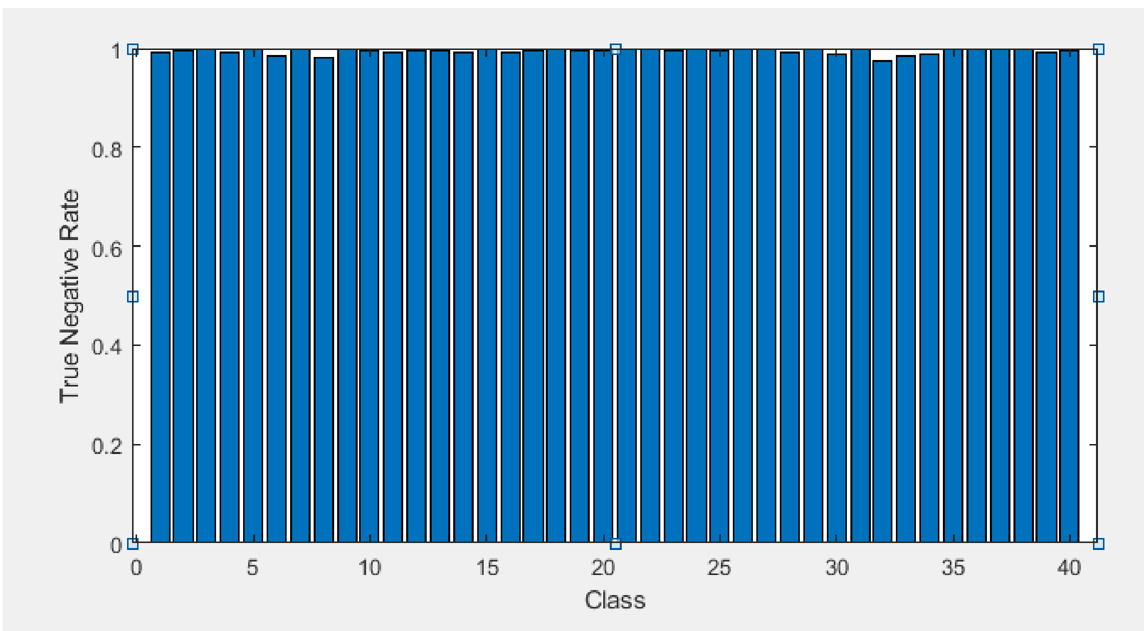


Figure 6.4: Specificity Rate for each class of KNN classifier

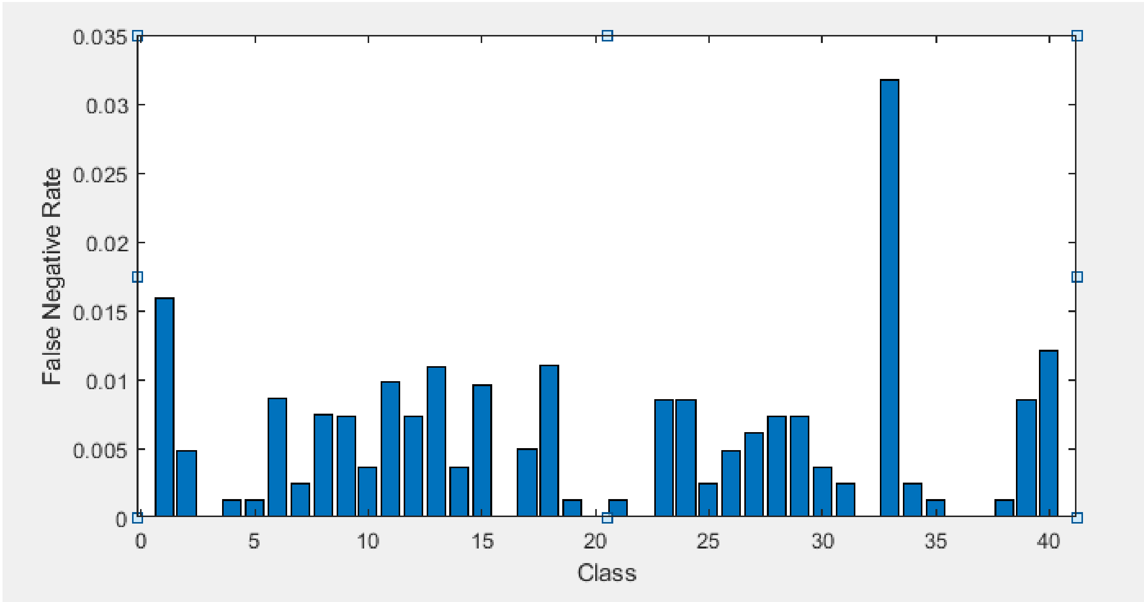


Figure 6.5: False Negative Rate for each class of KNN classifier.

6.3 Experimental Result & Performance Analysis on ORL Database

The ORL face database contains 400 face images of 40 different persons [29]. There are 10 images for each person. Those faces are centrally aligned. For some subjects, the images were taken at different lighting, facial expressions (open / closed eyes, smiling / not smiling) and facial details (glasses / no glasses). All the images were taken against a dark homogeneous background with the subjects in an upright, frontal position (with tolerance for some side movement). Among those 400 face images 280 images are used for training & 120 images are used for testing. In KNN classification k = 3 is used. PCA transforms face images onto Eigen faces. If the covariance matrix is of size n*n we will achieve n Eigen faces. But not all the Eigen faces carry relevant information. Some of the Eigen faces (the last ones) carry noise. So considering all the Eigen faces will make difficult for the classification algorithm. So, principal components are selected based on cumulative variance. As cumulative variance increases, the number of Principal Components also increases. So, by keeping the right trade-off between the energy maintenance and noise inclusion different numbers of principal component are selected. The performance of KNN classifier based on the selected number of components has been presented in following table.

Table 6.2: The performance of ORL Dataset

Cumulative variance	No. of Component	Test Accuracy on ORL (%)
<60	10	81.24
<65	14	82.13
<70	19	85.89
<75	23	87.92
<80	32	89.14
<85	43	88.67
<90	70	88.56

There are 40 different people i.e., 40 classes exist here. Maximum recognition rate using KNN classifier is 89.14%, when we select 30 components of training data set.

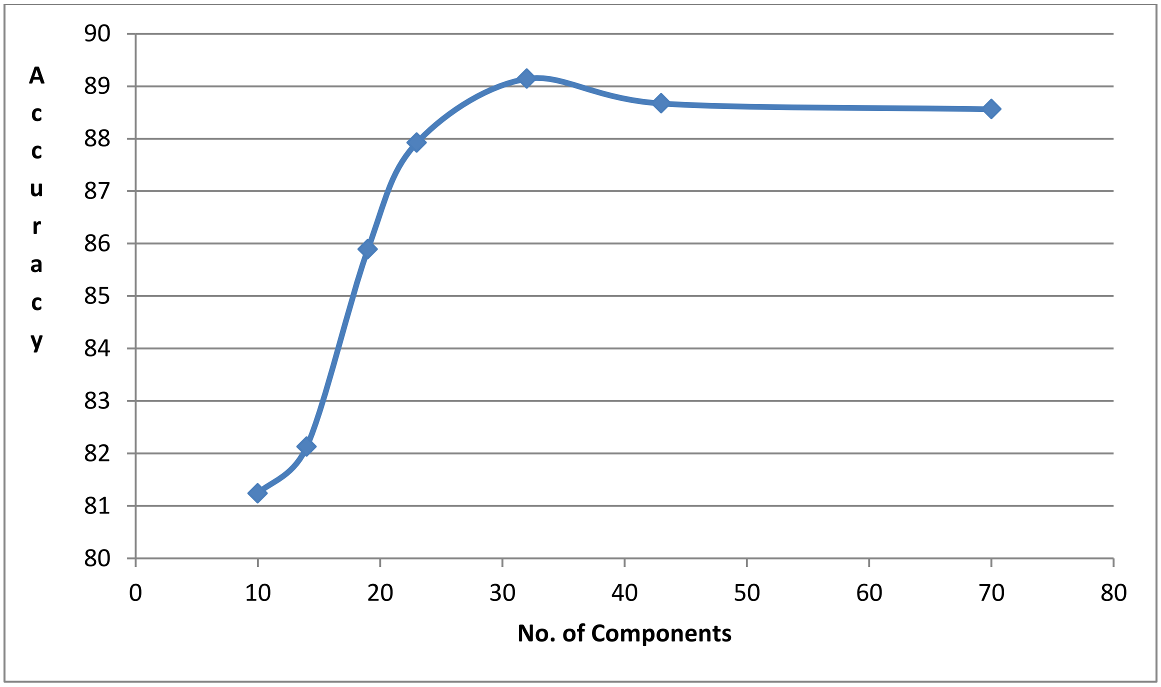


Figure 6.4: The performance of ORL Dataset using PCA & KNN classifier.

6.4 Conclusion

Viola-Jones face detector works well when images are frontal. But it's performance decreases when rotated images are taken into consideration. It can detect faces up to 40^o rotated images but includes some non-face area. Those areas cause the loss of performance when PCA feature extraction method is used for recognition purpose. Because PCA finds those axes where most variations exist, these non-face area cause irrelevant variation which results them fall away from the other objects of the same class.

Chapter 7
Conclusion

7.1 Introduction

In this thesis work an effort was made to detect faces by using Viola-Jones Face Detector & the detected portion is used for PCA feature extraction & classification is performed by KNN classifier. In this PCA face recognition approach face images are projected onto the feature space that best describes the variations among the face images. This approach transforms face image onto a small set of characteristic feature images called the "eigenfaces" which are the principal components of the initial training set of images. Recognition is performed by KNN classifier which is based on Euclidean distance measurement. The shortest K components are first selected & then the majority of the k item's classes are labeled as the predicted class level.

7.2 Limitations

Although the viola jones face detector offers real time performance, scale/location invariance, but still has a few limitations such as: it is not effective when it tries to detect rotated faces or tilted faces. There are some limitations of PCA face recognition also. It's recognition rate decreases when illumination & pose variation occur. This method is not robust when dealing with extreme change of expression. The faces need to be centrally aligned for better performance.

7.3 Future Works

In our thesis work, when rotated faces are considered for face detection, some non-face regions are included in the face image & faces are rotated. That's why our performance decreases. So if geometric alignment can be performed on face images, the performance of face recognition can improve. So, in future efforts can be made for geometric alignment for robust face recognition.

References

[1] Anil k. Jain, Patrick Flynn, Arun A. Ross, “Handbook of Biometrics”, *Springer Science & Business Media*, LLC, 2008.

[2] P. Viola, M. Jones, “Robust Real-Time Face Detection”, *International Journal of Computer Vision*, 57(2), 137-154, 2004.

[3] D. Parma and B. Mehta, “Face Recognition Methods & Applications”, *International Journal of Computer Technology & Applications*, vol. 4, no. 1, pp. 84-86, 2013.

[4] T. Sakai, M. Nagao, T. Kanade, “Computer analysis & classification of photographs of human faces", in proc. *First USA-Japan Computer Conference*, 1972, p. 2.7.

[5] W. Zhao, R. Chellappa, P. J. Phillips , A. Rosenfeld, “Face recognition: A literature survey”, *ACM Computing Surveys (CSUR)*, 35 (4) , 2013 399-458.

[6] Z. Kalal, K. Mikolajczyk, J. Matas, “Face-tld: Tracking-learning-detection applied to faces”, in: Image Processing (ICIP), 2010 17th *IEEE International Conference on, IEEE*, 2010, pp. 3789-3792.

[7] M. Pantic, L.J.M. Rothkrantz, “Automatic analysis of facial expressions: The state of the art”, *Pattern Analysis & Machine Intelligence, IEEE transactions*, 2000, pp. 1424-1445.

[8] N. kumer, A.C. Berg, P.N. Belhumear, S. K. Nayar, “Attribute & similar classifiers for face verification”, in Computer Vision, 2009 IEEE *12th International Conference on, IEEE*, 2009, pp. 365-372.

[9] Y. Fu, G. Guo, T. S. Huang, “Age synthesis & estimation via faces: A survey, Pattern Analysis & Machine Intelligence”, *IEEE Transactions*, 2010, pp. 1955-1976.

[10] P. Viola, M. Jones, “Rapid Object Detection using a Boosted Cascade of Simple Features”, *in proc. Of CVCR*, 2001.

[11] Y. Freund, R.E. Schapire, “A decision–theoretic genetic generalization of on-line learning and an application to boosting”, *Journal of Computer and System Sciences*, 1997, pp. 119-139.

[12] R.E. Schapire, Y. Singer, “Improved boosting algorithms using confidence-rated predictions”, *Machine learning*, 1999, pp. 297-336.

[13] B. Wu, H. Ai, C. Huang, S.Lao, “Fast rotation invariant multi-view face detection based on real adaboost”, *in:proc. Of IEEE Automatic face and Guesture recognition*, 2004.

[14] T. Mita, T. kaneko, O. Hori, “Joint Haar-like features for detection”, *in:proc. Of ICCV*, 2005.

[15] R. Lienhart, J.Maydt, “An extended set of Haar-like features for rapid object detection”, *in proc. Of ICIP*, 2002.

[16] S.C. Brubaker, J. Wu, J. Sun, M.D. Mullin, J.M. Rehg, “On the design of cascades of boosted ensembles for face detection”, *Tech. rep, Georgia Institute of Technology, GIT-GVU*, 2005.

[17] R. Xiao, H. Zhu, H. Sun, X. Tang, “Dynamic cascades for face detection”, *in proc of ICCV*, 2007.

[18] G. Zhang, X. Huang, S.Z. Li, Y. Wang , X. Wu, “Boosting local Binary pattern (LBP)- based face recognition”, *in proc, Advances in Biometric Authentication*, 2004.

[19] H. Jin, Q. Liu, H. Lu, X. Tong, “Face detection using improved lbp under Bayesian framework”, *in: third Intl. Conf. on Image & Graphics (ICIG)*, 2004.

[20] R. Baluga, M. Sahami, H. A. Rowley, “Efficient face orientation discrimination”, *in proc. Of ICIP*, 2004.

[21] K. Levi, Y. Weiss, “Learning object detection from a small number of examples: The importance of good featutres”, *in proc. Of CVPR*, 2004.

[22] Mrs. Sunita Roy, Prof. Samir K. Bandyopadhyay, “Face detection using a hybrid approach that combines HSV and RGB”, *IJCSMC*, vol. 2, issue 3, 2013, pp. 127-136.

[23] K. Schwerdt, J.L. Crowley, “Robust face tracking using color”, *AFGR00*, 2000.

[24] M. A. Turk and A. P. Pentland, “Eigenfaces for Face Detection /Recognition,” *Journal of Cognitive Neuroscience*, vol. 3, no. 1, pp. 1–11,1991.

[25] Erik Hjelmas, Boon Kee Low, Face Detection: A survey, Computer Vision & Image Understanding 83, 2001, pp. 236-274.

[26] Md Jan Nordin, Abdul Aziz K. Abdul Hamid, “Combining Local Binary Pattern and Principal Component Analysis on T-Zone Face Area for Face Recognition”, *International Conference on Pattern Analysis and Intelligent Robotics*, 2011.

[27] Ega Bima Putranto, Poldo Andreas Situmorang, Abba Suganda Girsang, “Face Recognition Using Eigenface with Naive Bayes”, *International Conference on Knowledge, Information and Creativity Support Systems (KICSS)*, Yogyakarta, Indonesia,2016,11.

[28] Datset: http://robotics.csie.ncku.edu.tw/Databases/ FaceDetect_PoseEstimate.htm

[29] Datset: https://www.cl.cam.ac.uk/research/dtg/attarchive/facedatabase.html.