LITERATURE SURVEY REPORT ON

Real-Time Emotion Detection from Face

Submitted in partial fulfillment of the requirements for the award of the degree of

Bachelor of Technology

In

Computer Science And Engineering

By

Manu Varghese

Reg No - 14004074

FEDERAL INSTITUTE OF SCIENCE AND TECHNOLOGY (FISAT)®

ANGAMALY-683577, ERNAKULAM (DIST)

Affiliated to

MAHATMA GANDHI UNIVERSITY

Kottayam-686560

November 2017

FEDERAL INSTITUTE OF SCIENCE AND TECHNOLOGY (FISAT)®

Mookkannor(P.O), Angamaly-683577

CERTIFICATE

This is to certify that Literature survey report for the project titled **Real-Time Emotion Detection from Face** is a bonafide work carried out by **Manu Varghese(Reg No-14004074)** in partial fulfilment for the award of Bachelor of Technology in Computer Science and Engineering from Mahatma Gandhi University, Kottayam, Kerala during the academic year 2017-2018.

Staff In-charge **Head of the Department**

Place:

Date:

ABSTRACT

The human face plays a prodigious role for automatic recognition of emotion in the field of identification of human emotion and the interaction between human and computer for some real application like driver state surveillance, personalized learning, health monitoring etc. Most reported facial emotion recognition systems, however, are not fully considered subject-independent dynamic features, so they are not robust enough for real life recognition tasks with subject (human face) variation, head movement and illumination change. For human-computer interaction facial expression makes a platform for non-verbal communication. The emotions are effectively changeable happenings that are evoked as a result of impelling force. So in real life application, detection of emotion is very challenging task. Facial expression recognition system requires to overcome the human face having multiple variability such as color, orientation, expression, posture and texture so on.A literature survey is done to investigate the various frameworks available for emotion detection from face.

ACKNOWLEDGEMENT

I would like to express my deepest appreciation to all those who have been instrumental in the successful completion of this work.

Mr.Paul Mundadan, chairman, FISAT Governing Body, who provided us with the vital facilities required.

Dr. George Issac, Principal, FISAT for the amenities he provided, which helped us in the fulfillment of this work.

Dr. Prasad J.C, HOD(CSE Dept), FISAT who always guided us and rendered his help in all ways possible.

Mr.Pankaj Kumar G, for his constant encouragement and enthusiastic supervision and for guiding us with patience in all the stages.Without his help and inspiration, this would not have been materialized.

Mrs. Divya John ,Mrs Resmi R ,Mrs. Preethi N P and Mr. Paul P Mathai and for their guidance and constant supervision as well as for providing necessary information and also for extending their support in completing this phase of project. The faculty of the CSE Dept., FISAT and Lab Instructors for providing us with the necessary Lab facilities.

My family who inspired, encouraged and fully supported us in every trial that came the way. Also, we thank them for giving us not just financial, but moral and spiritual support.

Contents

List of Figures i
List of Tables ii

1 Introduction **1**
1.1 Overview 1

2 Related Works **3**
2.1 A Robust Method for Face Recognition and Face Emotion Detection System using Support Vector Machines 3
2.2 A comparison of several Classifiers for Eye Detection on Emotion Expressing Faces 4
2.3 Facial-Feature Based Human-Computer Interface For Disabled People 5
2.4 Face Detection using Color Segmentation Thresholding 6
2.5 Support Vector Machine For Face Emotion Detection On Real Time Basis 7
2.6 Facial Emotion Recognition Based on Visual Information 8
2.7 Face recognition using Fisherface algorithm and elastic graph matching 9
2.8 An Efficient Method to Face and Emotion Detection 9
2.8.1 Feature Extraction 10
2.8.2 MEL Frequency Ceptral Coefficients 12
2.8.3 KNN Classifier 13
2.8.4 Proposed System 14
2.9 Facial Emotion Recognition in Real Time 17
2.9.1 Convolutional Neural Network 17
2.9.2 Using CNN 18
2.9.3 Experimental Results 20

3 Scope of the work **23**

4 Conclusion **25**

APPENDIX **27**

A Questions and Answers **28**

List of Figures

2.1 Integral Image 10
2.2 Sum Calculation 11
2.3 Different type of features 11
2.4 Database Creation for face Features 14
2.5 Database Creation for speech features 15
2.6 Feature Evaluation for Input Face 16
2.7 Feature evaluation for input voice 16
2.8 JAFFE Dataset 19
2.9 Extended Cohn-Kanade 19
2.10 Architecture of the convolutional neural network 20
2.11 Output of the second fully-connected layer 21
2.12 Emotion Prediction 22

List of Tables

2.1 Result Evaluation 17
2.2 Comparision of evolution result with different methods 17

Chapter 1

Introduction

Recently, diverse type of robots designed not only for industrial activities but also for communicating with human have been researched and developed. Therefore, opportunities for human who rarely use computers to have contact with robots are increasing. In addition, communication robots are aimed at accomplishing smooth communication with human. For these reasons, communication robots need more intuitive interaction systems. It is reasonable to suppose that robots that give human a sense of familiarity can communicate more smoothly with human. According to reports in the field of human communication, human feel familiarity with ones counterpart as they synchronize with their nonverbal information, e.g., facial expressions and voices. Robots can communicate more smoothly with human as they detect human emotions and respond with appropriate behaviors. Usually, almost all human express their own emotions with their facial expressions.The study of emotions has attracted interest of researchers from very diverse areas, ranging from psychology to the applied sciences. Face and emotion features detection is the currently very active area of research in the computer vision field as different kinds of face detection application are currently used such as image database management system, monitoring and surveillance analysis, biomedical image, smart rooms intelligent robots, human computer interfaces and drivers alertness system [1].

1.1 Overview

In last few decades very big amount of research is done in the field of face and emotion features detection.Face recognition and emotion recognition has got more attention and many advance technologies and methods are introduced for this in this period. Many commercial systems for face recognition are now available. New databases have been created and evaluations of recognition techniques using these databases have been carried out. Now, the face recognition has become one of the most active applications of pattern recognition, image analysis and understanding. Facial recognition plays a vital role in human computer interaction [2]. A Face recognition system can be either verification or an identification system depending on the context of an application. The verification system authenticates a person identity by

comparing the captured image with his/her own templates stored in the system. It performs a one to one comparison to determine whether the person presenting himself/herself to the system is the person he/she claims to be. An identification system recognizes a person by checking the entire template database for a match. It involves a one to many searches. The system will either make a match or subsequently identify the person or it will fail to make a match.

Generally three important steps involving in the face recognition system are [3]: (1) detection and rough normalization of faces, (2) feature extraction and accurate normalization of faces, (3) identification and/or verification. Face detection determines the location of human faces in an input image, which plays important roles in applications such as video surveillance, human computer interface, video conference and biometric application [4]. Automatic human face detection from images is a challenging task due to the variances in the image background, view, illumination, articulation, and facial expression. Based on the computer vision research, Haar wavelet is used for image feature detection for object recognition [5]. The success of the realtime face recognition systems are limited by the varying quality of images due to unreliable environment conditions. Hence, solutions to this problem are an active area of research and development. This is an effective method to create face and emotion feature database for the emotion and face detection of the human from voice and facial image respectively and then the database is going to use for evaluation of the human face and emotion.

Chapter 2

Related Works

2.1 A Robust Method for Face Recognition and Face Emotion Detection System using Support Vector Machines

[18]This research presents framework for real time face recognition and face emotion detection system based on facial features and their actions. The key elements of Face are considered for prediction of face emotions and the user. The variations in each facial feature are used to determine the different emotions of face. Machine learning algorithms are used for recognition and classification of different classes of face emotions by training of different set of images. In this context, by implementing algorithms would contribute in several areas of identification, psychological researches and many real world problems. The proposed algorithm is implemented using open source computer vision (OpenCV) and Machine learning with python.The face recognition is the basic part in modern authentication/identification applications; the accuracy of this system should be high for better results. Fisherface [20] algorithm presents high accurate approach for face recognition; it performs two classes of analyses to achieve recognition i.e. principal component analysis (PCA) and linear discriminant analysis (LDA) respectively. While dealing with the machine learning problems, dimensionality is the biggest issue. Therefore PCA is used to reduce the dimensionality of the images/frames. It converts the high dimensional space into low dimensional space. By reducing the dimensions, the number of features per image also be reduced. LDA is a discriminant method used in many recognition problems; it computes the group of characteristic features that normalizes the different classes if image data for classification. Fisherface is the best algorithm among others by its accuracy of around 96%. Detection of face in an image or video is the fundamental step in any recognition system.Face emotion recognition uses support vector machine for finding the different emotions of face and also for classifying them. PCA is used to extract the facial features and to reduce the image dimensions. Face is a two dimensional image, for face analysis it is preferred to use two dimensional vector space. Therefore for dimensionality reduction also 2DPCA is best for faces under different poses. 2DPCA is used to remove the unnecessary parts of the image. Multiobjective algorithm based

optimization and classifiers are used. SVMs are used to classify the image data under consideration. It finds the minimum possible separation between two or more classes of data and creates a hyper-plane with a margin. Since in general greater the margin and lesser is the generalization error. SVMs are memory efficient and effective in higher dimension spaces.

Face recognition which is implemented in real-time helps to recognize the human faces can be used for person identification and authentication purposes. Face emotion detection is implemented using support vector machine classifiers which are capable of classifying different class of emotions accurately. The accuracy of both face recognition and emotion detection can be increased by increasing the number of images during training. The detection time is significantly less and hence the system yields less run-time along with high accuracy. The implementation of the system in android improves the availability of the system to more users.

2.2 A comparison of several Classifiers for Eye Detection on Emotion Expressing Faces

[21]Eye Detection is an essential component in many image processing applications face recognition algorithms or human-computer interfaces rely on the accurate determination of the position of the eyes. For example, a computer security system will need to identify faces and this can be done by comparing them to references from a database. For a proper comparison, the faces must be aligned. As the position of the eyes and the interocular distance only slightly vary from one person to another, it means that the position of the eyes can be used to normalize and align a set of faces. There are other possible applications in the field of human- computer interaction the eyes are the key elements for the recognition and the classification of human emotions [11]. If computer can understand the users emotions, then a software program can adapt the communication to match the users state of mind. A potential application would be a teaching environment which knows when to ask easier questions or when to provide the user with some form of gratification. Alternatively, it can warn the user when an answer was given under a strong emotional state. Such systems are also useful to deal with medical conditions, such as autism in children. Psychological studies identified six types of facial expressions which are generally recognized. These are: fear, disgust, anger, happiness, sadness and surprise. When a person enters into one of these states, the face will change appearance significantly. In some cases, the eyes may become partially or fully closed, making the iris less visible. In other cases, such as when the person is surprised, the eyes might become much wider, producing a larger shining glint. All these changes of the eye appearance and of the other facial features (lowering eyebrows for example) will make precise eye localization more difficult, compared to the case of neutral state of the face.

This work gives a comparative study of three solutions to this non-trivial problem of finding the eyes on faces expressing emotions. It is based on machine learning techniques i.e. training a set of classifiers

so that they can distinguish between eye and non- eye regions on the face. Firstly, the classifiers will be trained with a set of positive and negative samples, followed by an evaluation of their performance. Several iterations will be made in order to find the best combination of classifier parameters and training environment. The algorithm is robust and works under various illumination conditions. The sole problems that might occur are in cases when very strong shadows appear on the eye region. In these extreme situations, the shape features of the eyes can be significantly modified, thus leading to errors. The eye detection process begins by scanning rectangular regions of 71x71 pixels from the image containing the face and classifying them into two categories: eye and non-eye. For each tested region in the image, this will compute a set of features, which will be used as input to the classifiers. The output of the classification engines will be further post- processed in order to improve the accuracy and determine the rectangular regions most likely to correspond to the eyes.

This section introduces the algorithms to compute feature vectors (constructed from image data) used as inputs for the classifiers .The image processing methods used for improving the accuracy of the final results are presented. The methodology used for evaluation is detailed. The goal was to train the classifiers to distinguish between eye and non-eye regions on a face expressing an emotion. As classifier input vectors, system used a integral projections encoded using the TESPAR method. The measurements revealed that each classification method has its own outstanding advantages, especially when it is tuned to our particular case. Moreover, in order to further enhance the results,a post-processing step is applied on the output of the classifiers. With this approach, it reaches detection ratios of over 95% while keeping the rate of incorrect decisions below 5%.The problem of detecting eyes in images with human emotions can be resolved by employing neural network classifiers. While the SVM provides the best accuracy, its computational demands make it less appealing for real-time systems for embedded use the Bayes or the MLP algorithms are more appropriate, especially when accompanied by carefully tuned post-processing step.Several iterations will be made in order to find the best combination of classifier parameters and training environment. The algorithm is robust and works under various illumination conditions. The sole problems that might occur are in cases when very strong shadows appear on the eye region. In these extreme situations, the shape features of the eyes can be significantly modified, thus leading to errors. The goal was to train the classifiers to distinguish between eye and non-eye regions on a face expressing an emotion.

2.3 Facial-Feature Based Human-Computer Interface For Disabled People

[1]Human-Computer Interaction Systems allowing for more natural communication with machines. Such systems are especially important for elderly and disabled persons. Face detection has always been a vast research field in the computer vision world, considering that it is the backbone of any application that

deals with the human face (e.g. biometric systems). The system presents a vision feature-based system for detection of long voluntary eye blinks and interpretation of blink patterns for communication between man and machine. Supplemented by the mechanism for detecting multiple eye blinks, this system provides a complete solution for building intelligent hands-free input devices. Due to recent increase of computer power and decrease of camera cost, it became very common to see a camera on top of a computer monitor. The described technique uses off-the self cameras that allow one for tracking nose features, eyebrows and head position robustly and precisely in both 2D and 3D coordinates. This tracking and monitoring allows user to give input to the computer machine and access the entire system in a hands free manner. The detailed theory behind the technology is presented in the paper and the results from running several perceptual user interfaces built with this technology are shown. HCI can be described as the point of communication between the human user and computer.

Typical input devices used nowadays for communication with the machine are: keyboard, mouse, trackball. Touch pad and a touch screen. All these interfaces require manual control and cant be used by person impaired in movement capacity. This fact induces the need for development of the alternative method of communication between human and computer that would be suitable for the disabled. Therefore the work on the development of innovative HCI attracts so much the attention of researchers all over the world. For severely paralyzed persons, whose ability of movement is limited o the muscles around the eyes most suitable are systems controlled by eyeblinks since blinking is the last voluntary action the disabled person loses control of. Eye blinks can be classified into 3 types: voluntary, reflexive and spontaneous. Spontaneous eye blinks are those with no external stimuli specified and they are associated with the psycho-physiological state of the person.Voluntary eye blinks are results of the persons decision to blink and can be used as a method for communication. The eye-movement or eye blink controlled HCI systems are very useful for persons who cannot speak or use hands to communicate (hemiparesis, quadriplegia, ALS). The systems use techniques based mainly on infrared light reflection or electrooculography. The example of gaze-communication device is Visionboard system [22]. The infrared diodes located in corners of the monitor allow for detection and tracking of the users eyes employing the bright pupil effect. The system replaces the mouse and keyboard of a standard computer and provides access to many applications, such as writing messages, drawing, remote control, Internet browsing or e-mail. However, majority of users were not fully satisfied with this solution and suggested improvements.

2.4 Face Detection using Color Segmentation Thresholding

Color segmentation[19] is an effective process to separate skin from its background. The color segmentation process will be followed by energy thresholding. Face detection has been a fascinating problem for image processing researchers during the last decade because of many important applications such as video face recognition at airports and security check-points, digital image archiving, etc. we attempt to

detect faces in a digital image using various techniques such as skin color segmentation, morphological processing, template matching, We determined that the more complex classifiers did not work as well as expected due to the lack of large databases for training. Reasonable results were obtained with color segmentation, template matching at multiple scales, and clustering of correlation peaks we try to replicate on a computer that which human beings are able to do effortlessly every moment of their lives, detect the presence or absence of faces in their field of vision. The model will take three different color spaces into consideration namely HSV,RGB and YCbCr. Assuming that a person framed in any random photograph is not an attendee at the gathering or get-together, it can be assumed that the face is not white, green, red, or any unnatural color of that nature. While different ethnic groups have different levels of melanin and pigmentation, the range of colors that human facial skin takes on is clearly a subspace of the total color space. With the assumption of a typical photographic scenario, it would be clearly wise to take advantage of face-color correlations to limit our face search to areas of an input image that have at least the correct color components. The color segmentation process will be followed by energy thresholding Thresholding is the operation of converting a grayscale image into a binary image. Thresholding is a widely applied preprocessing step for image segmentation. Often the burden of segmentation is on the threshold operation, so that a properly thresholded image leads to better segmentation. There are mainly two types of thresholding techniques available: global and local. In the global thresholding technique a grayscale image is converted into a binary image based on an image intensity value called global threshold. All pixels having values greater than the global threshold values are marked as 1 and the remaining pixels are marked as 0. In local thresholding technique, typically a threshold surface is constructed that is a function on the image domain. propose to develop a model for face detection based n color segmentation. The color segmentation process will be followed by energy thresholding. The model tries to take advantage of face color correlation. The model will take three different color spaces into consideration namely HSV,RGB and YCB.

2.5 Support Vector Machine For Face Emotion Detection On Real Time Basis

Enabling computer systems to recognize facial expressions and infer emotions from them in real time presents a challenging research topic.A real-time method is proposed as a solution to the problem of facial expression classification in video sequences. The system employ an automatic facial feature tracker to perform face localization and feature extraction. The facial feature displacements in the video stream are used as input to a Support Vector Machine classifier[13]. It evaluates method in terms of recognition accuracy for a variety of interaction and classification scenarios. The person-dependent and person-independent experiments demonstrate the effectiveness of a support vector machine and feature tracking approach to fully automatic, unobtrusive expression recognition in live video.

Automatic face detection and recognition are two challenging problems in the domain of image processing and computer graphics that have yet to be perfected. Manual recognition is a very complicated task where it is vital to pay attention to primary components like: face configuration, orientation, location where the face is set (relative to the body), and movement (i.e. traces out a trajectory in space). It is more complicated to perform detection in real time. Dealing with real time capturing from a camera device, fast image processing would be needed. Haar features by Viola and Jones [8] is the first real time frontal-view face detector. Hence, we propose the use this method in this project.Facial recognition can be attempted once the face is detected in the image. There are a number of different approaches to performing face recognition, which have varying levels of success. Some of the better-known algorithms utilize eigenfaces [14] or active appearance models [15] to identify an image. However, eigenface approaches suffer from requiring extremely constrained frontal face images and potentially large amounts of training data to deal with high variability. Active appearance models are more promising for a noisy environment but require computationally expensive models. In geometric feature-based methods [16] facial features such as eyes, nose, mouth, and chin are detected. Properties and relations such as areas, distances, and angles, between the features are used as the descriptors of faces. This paper attempt at facial recognition follows the popular use of machine learning to determine the differences between features. Support Vector Machines (SVMs) have been recently proposed by C.Corinna[17] together with their co-workers as a very effective method for general purpose pattern recognition. By utilizing Support Vector Machines (SVM) to create models, it enables the creation of a complicated description of what features characteristics determine an expression.

2.6 Facial Emotion Recognition Based on Visual Information

Facial Emotion Recognition (FER)[23] is an important topic in the fields of computer vision and artificial intelligence owing to its significant academic and commercial potential. Although FER can be conducted using multiple sensors, this review focuses on studies that exclusively use facial images, because visual expressions are one of the main information channels in interpersonal communication. It provides a brief review of researches in the field of FER conducted over the past decades.

First, conventional FER approaches are described along with a summary of the representative categories of FER systems and their main algorithms. Deep-learning-based FER approaches using deep networks enabling learning are then presented. This review also focuses on an up- to-date hybrid deeplearning approach combining a Convolutional Neural Network (CNN) for the spatial features of an individual frame and Long Short-Term Memory (LSTM) for temporal features of consecutive frames.

A brief review of publicly available evaluation metrics is given, and a comparison with benchmark results, which are a standard for a quantitative comparison of FER researches, is described. This review can serve as a brief guidebook to newcomers in the field of FER, providing basic knowledge and a general

understanding of the latest state-of-the-art studies, as well as to experienced researchers looking for productive directions for future work.

2.7 Face recognition using Fisherface algorithm and elastic graph matching

This System proposes a face recognition technique that effectively combines elastic graph matching[20] (EGM) and the Fisherface algorithm. EGM as one of the dynamic link architectures uses not only faceshape but also the gray information of image, and the Fisherface algorithm as a class-specific method is robust about variations such as lighting direction and facial expression. In the proposed system it mainly adopt the above two methods, the linear projection per node of an image graph reduces the dimensionality of labeled graph vector and provides a feature space to be used effectively for the classification. In comparison with the conventional method, the proposed approach could obtain satisfactory results from the perspectives of recognition rates and speeds. In particular, could get maximum recognition rate of 99.3% by the leaving-one-out method for experiments with the Yale face databases. As Information Age develops, the security of information is becoming more and more important and access to a reliable personal identification is becoming increasingly essential. Because conventional methods of identification based on possessiori of ID card or exclusive knowledge like a social security number or a password are not altogether reliable, biometrics that make out ones identity and authentication can be used. Especially, in the perspective of ease of use and accuracy, face recognition has an advantage compared with other biometrics. In general face recognition procedure, the most important thing is which feature vector is used. In early stage, face recognition by using Karhunen-Loeve (K-L) projection was proposed. And several methods such as Fisherface and elastic graph matching (EGM) were researched. Principal component analysis (PCA) and Fisherface using K-L projection are used to reduce the dimensionality of the feature vector and classify the feature space. But these methods have a defect that recognition rate decreases rapidly as the transition of a face region happens. In the case of EGM, that problem can be solved by Global Move and its recognition rate is higher than the above methods also. But compared with methods using K-L projection, its recognition speed is so slow that the recognition

2.8 An Efficient Method to Face and Emotion Detection

Most of the face recognition (FR) approaches have focused on the use of two dimensional images. Since FR is still an unsolved problem under the different conditions, such as pose, illumination or database size. The expression of emotions and the recognition of a person affective state are abilities indispensable for natural human interaction and social integration. The study of emotions has attracted interest of researchers from very diverse areas, ranging from psychology to the applied sciences. Face and emotion features detection[12] is the currently very active area of research in the computer vision field as different kinds

of face detection application are currently used such as image database management system, monitoring and surveillance analysis, biomedical image, smart rooms intelligent robots, human computer interfaces and drivers alertness system

2.8.1 Feature Extraction

VIOLA-JONES FACE DETECTION

Viola Jones is the oldest and most recognized face algorithm available for the face detection from the image. The basic principle of the Viola-Jones algorithm is to scan a sub-window capable of detecting faces across a given input image. The standard image processing approach would be to rescale the input image to different sizes and then run the fixed size detector through these images. This approach turns out to be rather time consuming due to the calculation of the different size images. Contrary to the standard approach Viola-Jones[8] rescale the detector instead of the input image and run the detector many times through the image each time with a different size. At first one might suspect both approaches to be equally time consuming, but Viola-Jones have devised a scale invariant detector that requires the same number of calculations whatever the size. This detector is constructed using a so-called integral image and some simple rectangular features reminiscent of Haar wavelets. The first step of the Viola-Jones face detection algorithm is to turn the input image into an integral image. This is done by making each pixel equal to the entire sum of all pixels above and to the left of the concerned pixel. This is demonstrated in Figure 2.1.

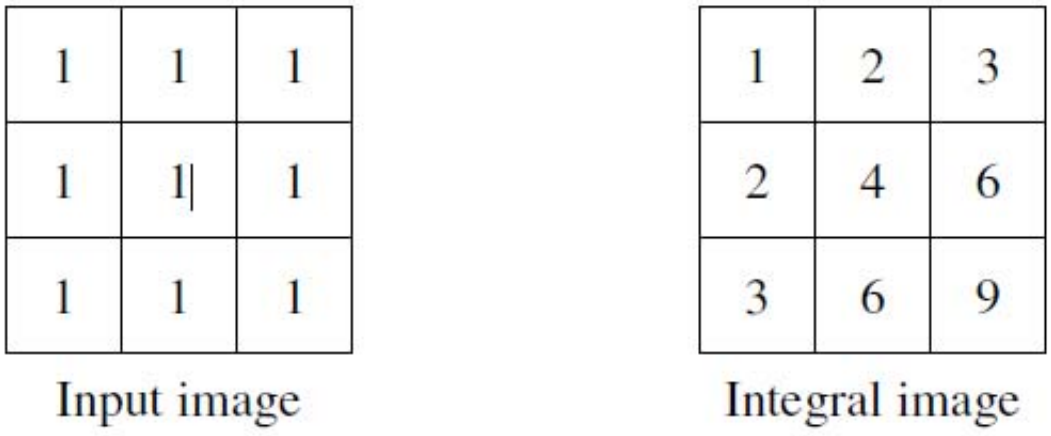

Figure 2.1. Integral Image

This allows for the calculation of the sum of all pixels inside any given rectangle using only four values. These values are the pixels in the integral image that coincide with the corners of the rectangle in the input image. This is demonstrated in Figure 2.2.

Since both rectangle B and C include rectangle A the sum of A has to be added to the calculation. It has now been demonstrated how the sum of pixels within rectangles of arbitrary size can be calculated

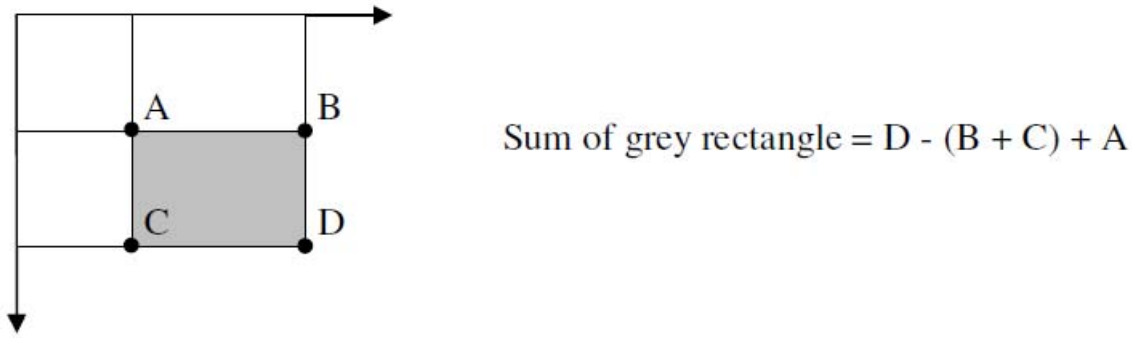

Figure 2.2. Sum Calculation

in constant time A human can do this easily, but a computer needs precise instructions and constraints. To make the task more manageable, Viola Jones requires full view frontal upright faces. Thus in order to be detected, the entire face must point towards the camera and should not be tilted to either side. While it seems these constraints could diminish the algorithm's utility somewhat, because the detection step is most often followed by a recognition step, in practice these limits on pose are quite acceptable.The Violan Jones face detector analyzes a given sub-window using features consisting of two or more rectangles. The different types of features are shown in figure 2.3.

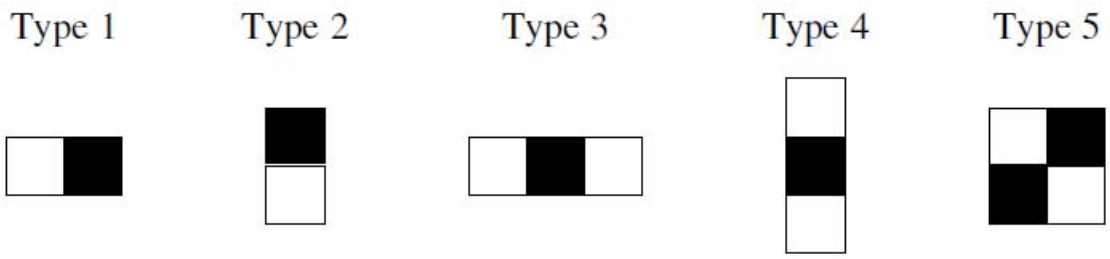

Figure 2.3. Different type of features

Each feature results in a single value which is calculated by subtracting the sum of the white rectangle(s) from the sum of the black rectangle(s). Viola-Jones has empirically found that a detector with a base resolution of 24*24 pixels gives satisfactory results. When allowing for all possible sizes and positions of the features in Figure 4 a total of approximately 160.000 different features can then be constructed. Thus, the amount of possible features vastly outnumbers the 576 pixels contained in the detector at base resolution.

These features may seem overly simple to perform such an advanced task as face detection, but what the features lack in complexity they most certainly have in computational efficiency. One could understand the features as the computers way of perceiving an input image. The hope being that some features will

yield large values when on top of a face. Of course operations could also be carried out directly on the raw pixels, but the variation due to different pose and individual characteristics would be expected to hamper this approach. As stated above there can be calculated approximately 160.000 feature values within a detector at base resolution. Among all these features some few are expected to give almost consistently high values when on top of a face. In order to find these features Viola-Jones use a modified version of the AdaBoost algorithm.An important part of the modified AdaBoost algorithm is the determination of the best feature, polarity and threshold. There seems to be no smart solution to this problem and 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. This is expected to be the most time consuming part of the training procedure. The best performing feature is chosen based on the weighted error it produces. This weighted error is a function of the weights belonging to the training examples.

2.8.2 MEL Frequency Ceptral Coefficients

MFCC are very useful features for audio processing in clean conditions. However, performance using MFCC features deteriorates in the presence of noise. There has been an increased effort in recent times to find new features that are more noise robust compared to MFCCs. Features such as, spectro-temporal modulation features [6] are more robust to noise but are computationally expensive. Skowronski and Harris [7] suggested modification of MFCC that uses the known relationship between center frequency and critical bandwidth. They also studied the effects of wider filter bandwidth on noise robustness. Herein, we suggest different modifications to MFCCs that make it more robust to noise without adding prohibitive computational costs. MFCC features approximate the frequency decomposition along the basilar membrane by a short-time Fourier Transform. The auditory critical bands are modeled using triangular filters, compression is expressed as a log function and a discrete cosine transform (DCT) is used to decorrelate the features.

Speech signals contain two types of information, time and frequency. In time space, sharp variations in signal amplitude are generally the most meaningful features. In the frequency domain, although the dominant frequency channels of speech signals are located in the middle frequency region, different speakers may have different responses in all frequency regions . Thus, the traditional methods which just consider fixed frequency channels may lose some useful information in the feature extraction process. The multi-resolution decomposing technique using wavelet transform is used to solve this problem. Based on this technique, one can decompose the speech signal into different resolution levels. The characteristic of multiple frequency channels and any change in the smoothness of the signal can then be detected to perfectly represent the signals. Then, the MFCCs are applied to the wavelet channels to extract features characteristics. MFCCs as previously stated, has the advantage that they can represent sound signals in an efficient way because of the frequency warping property. In this way, the advantages of both techniques

are combined in the proposed technique.

2.8.3 KNN Classifier

The k Nearest Neighbor (KNN) is one of the most commonly used methods for pattern recognition and has been applied in a variety of cases [5]. KNN Classifier works as follows. First for each one of the training set elements a classification of it is performed based on various neighborhoods. The k value that maximizes the DC of each classification is found. Therefore, for each training set there corresponds a particular k value which is considered the best available. Afterwards, for each unknown element, the nearest neighbor is found and its k value is assumed (based on the optimum k array). Then, the KNN classifier is applied on that test element, using that k value. As a concept, this is something similar to one of the ideas presented by Abidin and Perrizo [5]. K nearest neighbors search algorithm maintains a priority queue. The entries of the queue are Minimum Bounding Rectangles (MBRs) and objects which will be examined by the algorithm and are sorted according to their distance from the query point. An object will be examined when it reaches the top of the queue. The algorithm begins by inserting the root elements of the R-tree in the priority queue. Then, it selects the first entry and inserts its children. This procedure is repeated until the first data object reaches the top of the queue. This object is the first nearest neighbor. The KNN Classifier algorithm is shown below. Therefore, each object of the test set is a query point and each object of the training set, contains an additional attribute which indicates the class where the object belongs to. The R-tree is built using the objects of the training set.

Algorithm

1. PriorityQueue.enqueue(roots children)
2. NNCounter = 0
3. while PriorityQueue is not empty and NNCounter k do
4. element = PriorityQueue.dequeue()
5. if element is an object or its MBR then

. if element is the MBR of Object and PriorityQueue is not empty and objectDist(q, Object) > PriorityQueue.top then

7. PriorityQueue.enqueue(Object, ObjectDist(q, Object))
8. else
9. Report element as the next nearest object (save the class of the object)
10. NNCounter++
11. if early-break conditions are satisfied then
12. Classify the new object q in the class where the most nearest neighbors belong to and break the while loop. q is classified using NNCounter nearest neighbors
13. endif
14. endif

15. else if element is a leaf node then
16. for each entry (Object, MBR) in element do
17. PriorityQueue.enqueue (Object, dist(q, Object))
18. endfor
19. else /*non-leaf node*/
20. for each entry e in element do
21. PriorityQueue.enqueue(e, dist(q, e))
22. endfor
23. endif
24. end while
25. if no early-break has been performed then // use k nearest neighbors
26. Find the major class (class where the most nearest neighbors belong to)
27. Classify the new object q to the major class
28. endif

2.8.4 Proposed System

The main aim is to implement an efficient method to detect the face and emotion of the person. Work is divided into two parts for the storing the features of the face and the features of the voice of human and second evaluation of the face and emotion of the person using the features database.

Phase 1:

We use Viola Jones face detection algorithm to create the database for the face detection. The numbers of input images are collected for the creation of database. Figure 2.4 shows the overall implementation for the database creation for the face features detection and database creation.

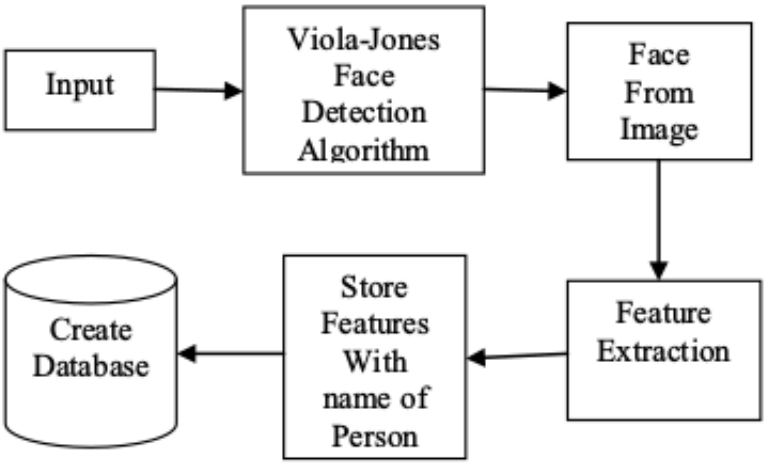

Figure 2.4. Database Creation for face Features

Input images are used to create the database, Viola Jones face detection algorithm is applied to the input image. After the algorithm applied the face from the image is detected and different features of the image are calculated from the face of image. To store in the database area of the face, edges of the face, major axis and minor axis from the face and shape type or eccentricity of the face, this type of features are calculated. If the eccentricity of the face is 0 the shape of the face is circle and if the eccentricity of the face is 1 then the shape of the face is going to be line the database. Once we calculated the all this features from the face this features are store in the database with the name of the person.

Different Mel frequency components of the voice input are calculated with the algorithm. There are total 22 Mel frequency components of the voice are calculated. Now from the each Mel component of the features are calculated. There are total 281 different features are calculate from the Mel component. Ones all the features are calculated this features are store in the database with the name of the person and the emotion type of the voice. For creating the database we are using following type of the emotion type sample: Happy, Sad, Angry, Surprise and Normal.

Phase 2:

In phase 2 KNN classifier algorithms is used to find face and emotion of the person. In evaluation phase input image and voice of the person give as the input to the system. Features of the face are calculated by using Viola Jones face detection algorithm. Mainly five features of the image are calculated from the input image i.e. area of the face, edges of the face, major axis and minor axis from the face and shape type or eccentricity of the face. Now the KNN algorithm is used to evaluate the similar features from the database created in the phase 1. KNN classifier algorithm gives the name of the person whose features are matched from the database.Figure 2.5shows the overall implementation for the database creationfor the speech features detection and database creation.

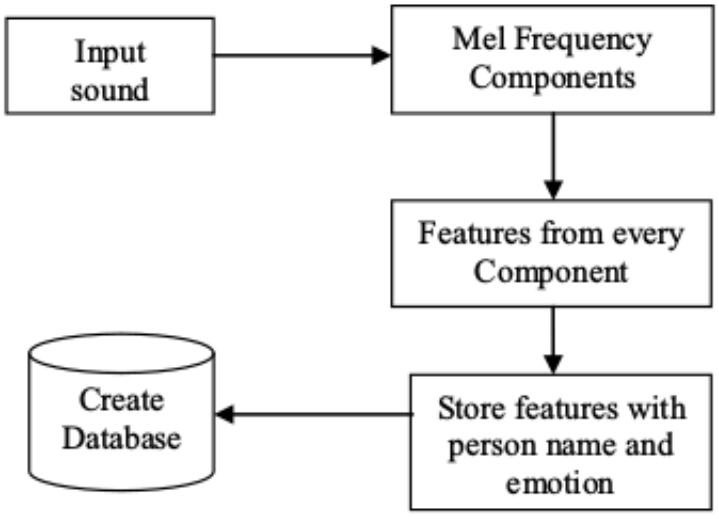

Figure 2.5. Database Creation for speech features

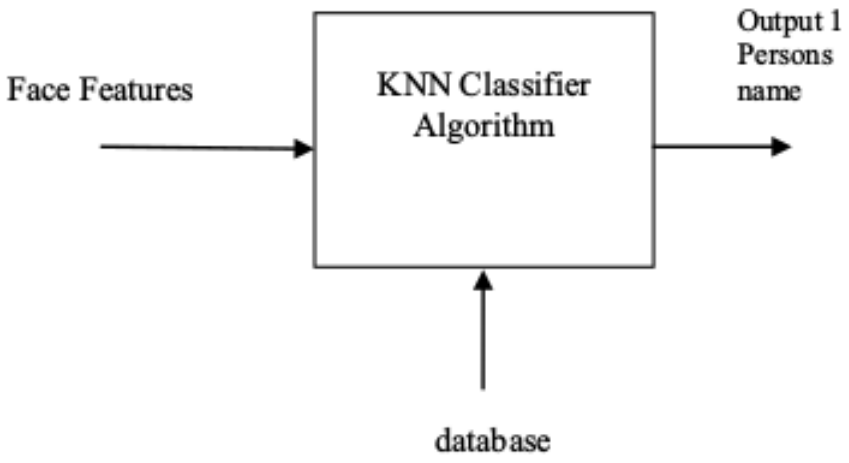

Figure 2.6. Feature Evaluation for Input Face

Now person voice gives as the input to the system to find out the emotion of the person. Different voice features of the input voice are calculated with the help of Mel frequency components of the voice. MFCC, base frequency, pitch and harmonic features of the voice are considered for the database creation and evaluation phase. Now the KNN algorithm is used to evaluate the similar features from the database created in the phase 1 as in figure 2.6. KKN classifier algorithm gives the emotion of the person whose features are matched from the database.

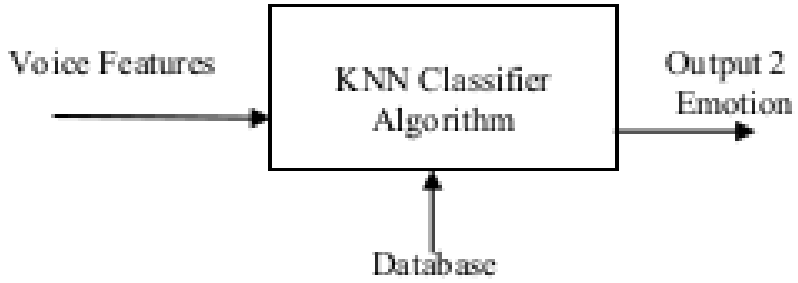

Figure 2.7. Feature evaluation for input voice

Now input 1 and input 2 from face and emotion reorganization is shown in figure 2.7,it gives complete output of the proposed system.

The table 2.1 shows the Experiment results by the proposed methodology.The obtained results are Compared with the other methods and it is shown in table 2.2.

No. face features in DB	No. voice features in DB	Accuracy of Face detection	Accuracy of emotion detection	Average Accuracy
10	10	94%	95%	94.5%
20	20	94.6%	95.2%	94.9%
30	30	96%	96.5%	96.25%
40	40	96.8%	97%	96.9%

Table 2.1. Result Evaluation

Different Algorithms	Face Detection Accuracy	Emotion Detection Accuracy
Proposed Work	96 %	95.7 %
Mingyu You. [17]	-----	72 %
Viola-Jones [18]	95.5 %	-----
MLLL [18]	96 %	-----
MLLL+BILBO 15 landmarks [18]	94 %	-----

Table 2.2. Comparision of evolution result with different methods

2.9 Facial Emotion Recognition in Real Time

Emotions often mediate and facilitate interactions among human beings. Thus, understanding emotion often brings context to seemingly bizarre and/or complex social communication. Emotion can be recognized through a variety of means such as voice intonation, body language, and more complex methods such electroencephalography (EEG) [10]. However, the easier, more practical method is to examine facial expressions. There are seven types of human emotions shown to be universally recognizable across different cultures [11]: anger, disgust, fear, happiness, sadness, surprise, contempt. Interestingly, even for complex expressions where a mixture of emotions could be used as descriptors, cross-cultural agreement is still observed.Therefore a utility that detects emotion from facial expressions would be widely applicable. Such an advancement could bring applications in medicine, marketing and enter- tainment.

2.9.1 Convolutional Neural Network

In machine learning, a convolutional neural network (CNN, or ConvNet) is a class of deep, feed-forward artificial neural networks that has successfully been applied to analyzing visual imagery. CNNs use a variation of multilayer perceptrons designed to require minimal preprocessing. They are also known as shift invariant or space invariant artificial neural networks (SIANN), based on their shared-weights architecture and translation invariance characteristics. Convolutional networks were inspired by biological

processes in that the connectivity pattern between neurons resembles the organization of the animal visual cortex. Individual cortical neurons respond to stimuli only in a restricted region of the visual field known as the receptive field. The receptive fields of different neurons partially overlap such that they cover the entire visual field.Also, such network architecture does not take into account the spatial structure of data, treating input pixels which are far apart in the same way as pixels that are close together. Thus, full connectivity of neurons is wasteful for purposes such as image recognition that are dominated by spatially local input patterns. As opposed to MLPs, CNNs have the following distinguishing features: 3D volumes of neurons.

The layers of a CNN have neurons arranged in 3 dimensions: width, height and depth. The neurons inside a layer are connected to only a small region of the layer before it, called a receptive field. Distinct types of layers, both locally and completely connected, are stacked to form a CNN architecture. Local connectivity: following the concept of receptive fields, CNNs exploit spatial locality by enforcing a local connectivity pattern between neurons of adjacent layers. The architecture thus ensures that the learnt "filters" produce the strongest response to a spatially local input pattern. Stacking many such layers leads to non-linear filters that become increasingly global (i.e. responsive to a larger region of pixel space) so that the network first creates representations of small parts of the input, then from them assembles representations of larger areas. Shared weights: In CNNs, each filter is replicated across the entire visual field. These replicated units share the same parameterization (weight vector and bias) and form a feature map. This means that all the neurons in a given convolutional layer respond to the same feature within their specific response field. Replicating units in this way allows for features to be detected regardless of their position in the visual field, thus constituting the property of translation invariance. Together, these properties allow CNNs to achieve better generalization on vision problems. Weight sharing dramatically reduces the number of free parameters learned, thus lowering the memory requirements for running the network and allowing the training of larger, more powerful networks.

2.9.2 Using CNN

1.Dataset

To develop a working model, two different freely- available datasets:

1)Extended Cohn-Kanade dataset (CK+)

2) the Japanese Female Facial Expression (JAFFE) database.

The CK+ dataset, although small, provides well-defined facial expressions in a controlled laboratory environment. The JAFFE database provides additional images with more subtle facial expressions with laboratory conditions.The sample image in JAFFE Dataset is shown in figure 2.8. This uniquely developed our own new (home-brewed) database that consists of images from five individuals. Many images were recorded for each of the seven primary emotions (anger, disgust, fear, happy, neutral, sad, surprise) from each subject. System subsequently applied Filter to these images to account for variations in

lighting and subject position in the final implementation Initially we directly implemented the VGG S

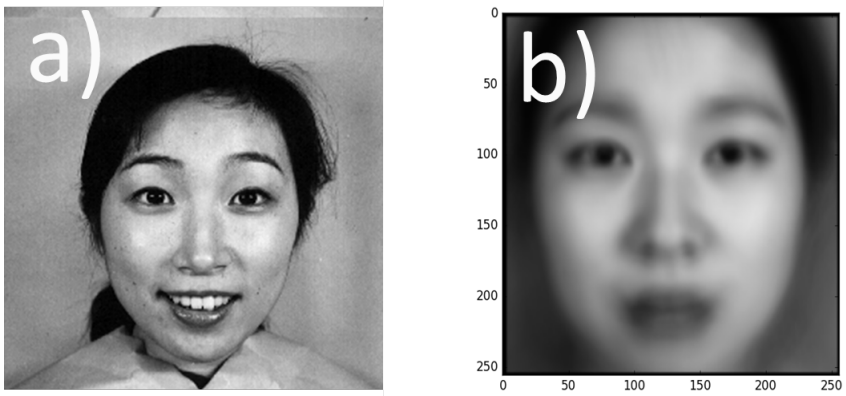

Figure 2.8. JAFFE Dataset

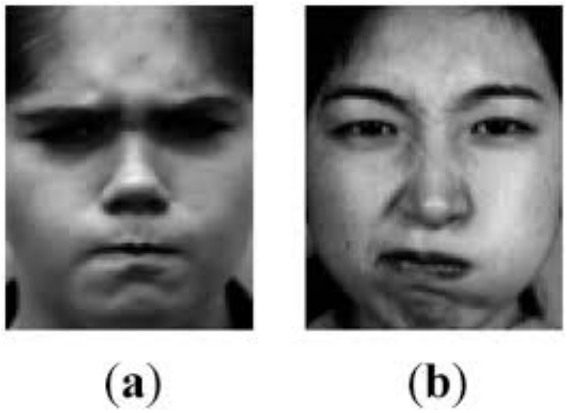

Figure 2.9. Extended Cohn-Kanade

network[11] for image classification. We were unable to obtain similar results and at best could obtain a test accuracy of 24% on the CK+ dataset, and 14% on the JAFFE dataset. There were incorrect classifications for images that appeared to clearly convey a particular emotion, such as that in Figure 2.9(a). The system pre-processed the data by subtracting the mean image (Figure 2.9(b)), but the results showed no improvement. Many different facial expressions were incorrectly classified as fear by VGG S for both datasets.It is not clear why the accuracy was lower. One issue is that the facial expressions in the JAFFE dataset are quite subtle, exacerbating the ability to differentiate emotions. Another issue is that there are few images labeled with fear and disgust in both the JAFFE and CK+ datasets, making it difficult to train the network to recognize these two emotions correctly. To improve the classification accuracy of VGG S, we applied transfer learning to the network on the JAFFE and CK+ datasets, as well as our own dataset. For our own home-brewed dataset, we recorded images from 5 different individuals. Multiple images were recorded for each of the seven primary emotions (anger, disgust, fear,

happy, neutral, sad, surprise) from each subject, resulting in a total of 2118 labeled images.The emotions disgust and contempt are avoided,since they are very difficult to classify. In any implementation of a CNN in a real environment, effects that are usually omitted in a laboratory must be accounted for. These include variations in lighting, distance from the camera, incorrect face cropping, and variations in orientation of the subjects face. A randomized Jitter implemented on the the home-brewed dataset. The jitter was applied by randomly changing both cropping (10% variation) and bright- ness (20% variation) in the input images. By Re-examining the same images with different variations in cropping and lighting, VGG S could learn to account for these effects.

Due to time constraints, It only trained the last few fully-connected layers of the network (fc6, fc7, fc8). The architecture of VGG S consists of five convolutional layers, three fully-connected layers, followed by a softmax classifier, for a total of approximately 7x10 6 convolutional parameters and 7x10 7 fully-connected parameters. The system applied a Haar-Cascade filter provided by OpenCV to crop the input image faces, which significantly improved test and training accuracy.A small batch size was required for training due to lengthy computation times and is the cause of the jagged profile of the curve. Convergence to low loss appears to occur after 100 iterations.

2.9.3 Experimental Results

Once a newly trained version of VGG S was obtained, this connected a video stream to the network using a standard webcam. The run-time for image cropping using the face-detector was 150 ms and that for a forward pass in VGG S was 200 ms. The Figure 2.10 shows the architecture of convolutional neural networks. These operations limited the frame-rate of our emotion-recognition algorithm to 2.5

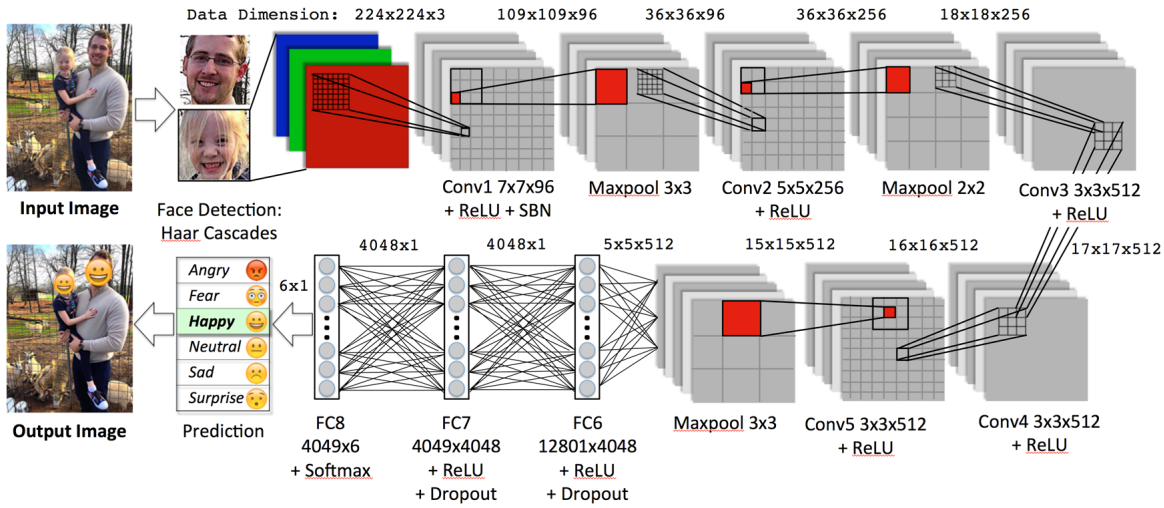

Figure 2.10. Architecture of the convolutional neural network

frames/second, sufficient for a real-time demonstration. Note that for any number N of subjects in the cameras view, the run-time for a single frame would be increased to 150 ms + N x 200 ms to account for

all persons in the frame. The network was developed on a laptop that had insufficient GPU capabilities (for VGG S) to utilize CUDNN,a GPU-accelerated library provided by Nvidia for deep neural networks. Thus, other machines with optimized graphics capabilities, such as the Jetson TX1, could implement our solution with substantially reduced run-time. We tried to implement VGG S on the Jetson TK1, but found that the run-time memory requirements ($>$ 2 GB) were too large for the board to process without crashing. It is interesting to examine the second fully-connected layers neurons responses to different facial expressions.

The system use the cosine similarity, to measure the similarity between two outputs A and B from the second layer (fc7). Figure 2.11 shows the output of the second fully-connected layer for two different input images displaying the emotion happy. The cosine similarity be- tween each of the layers in the output is quite high, revealing an average similarity of s = 0.94.

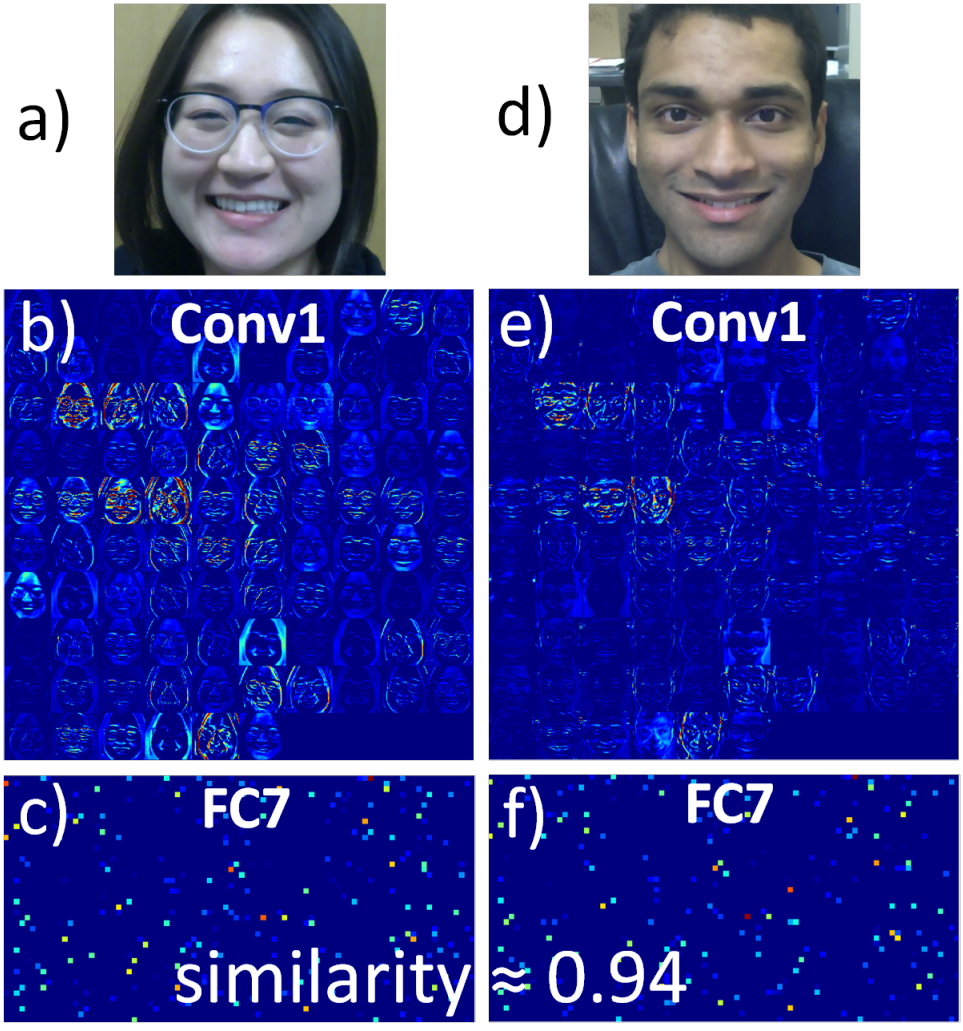

Figure 2.11. Output of the second fully-connected layer

Figure 2.12. Emotion Prediction

The Result is shown in the figure 2.12.The Emotions are shown using corresponding Emoji's. Augmenting our home-brewed dataset to include off-center faces could have addressed this problem. Heavier pre-processing of the data would have certainly improved test-time accuracy.Adjusting the brightness to the same level on all the images might have removed the requirement for providing jittered input images. Also, fully training a network other than VGG S might yield substantial improvements in computa- tion speed, since VGG S is relatively slow.

Chapter 3

Scope of the work

Robots that communicate with human have attracted much attention in the research field of robotics. In communication between human, almost all human recognize the subtleties of emotion in each others facial expressions, voices, and motions. Robots can communicate more smoothly with human as they detect human emotions and respond with appropriate behaviors. Usually, almost all human express their own emotions with their facial expressions.In actual communication, it is possible that some parts of the face will be occluded by adornments such as glasses or a hat. In previous studies on facial recognition, these studies have been had the process to fill in the gaps of occluded features after capturing facial features from each image. However, not all occluded features can always be filled in the gaps accurately. Therefore, it is difficult for robots to detect emotions accurately in real-time communication. For this reason,an emotion detection system is taking into consideration,partial occlusion of the face using causal relations between facial features.

Video games belong to the wide area of entertainment applications. Thus, assuming the existence of human emotions and in fact basing on them, they attempt to make the player to become emotionally attached with them. As the primary goal of a video game is to entertain the player, each video game try to allow the player to fulfill his or her dream. Standard video games try to do it in different ways depending on their genre and involving such elements as good gameplay, immersing storytelling, novelty, graphics and so on. Although video games belong to applications in which emotions naturally play an important role, only few of them try to incorporate their players affective state into the gameplay. Such games can be referred as affective or more properly affect-aware games. The importance of affect in delivering engaging experiences in entertainment and educational games is well recognized. Potential profits for affect-aware video games are not to be underestimated. Unfortunately, this affect-awareness is usually statically built-in the game at its development stage basing on the assumed model of so called representative player. There are two problems with such attitude. Firstly, each player differs in some way from that averaged model. Secondly, and more important, players affect state can change even radically

from session to session making almost impossible to predict the current user emotions at the development stage

Eye Detection is an essential component in many image processing applications face recognition algorithms or human computer interfaces rely on the accurate determination of the position of the eyes. For example, a computer security system will need to identify faces and this can be done by comparing them to references from a database. For a proper comparison, the faces must be aligned. As the position of the eyes and the interocular distance only slightly vary from one person to another, it means that the position of the eyes can be used to normalize and align a set of faces. There are other possible applications in the field of human computer interaction the eyes are the key elements for the recognition and the classification of human emotions. If we can make a computer understand the users emotions, then a software program can adapt the communication to match the users state of mind. A potential application would be a teaching environment which knows when to ask easier questions or when to provide the user with some form of gratification. Alternatively, it can warn the user when an answer was given under a strong emotional state. Such systems are also useful to deal with medical conditions, such as autism in children.

Chapter 4

Conclusion

Face recognition which is implemented in real-time helps to recognize the human faces can be used for person identification and authentication purposes.Various Methods for implementing the same is studied.Most reported facial emotion recognition systems, however, are not fully considered subject independent dynamic features, so they are not robust enough for real life recognition tasks with subject (human face) variation, head movement and illumination change.

The feature extraction and emotion prediction using the classifiers like KNN and SVM are found to be time consuming.It need the face features and voice features for the emotion prediction.The accuracy of both face recognition and emotion detection can be increased by increasing the number of images during training.From this survey it has been understood that extracting the feature of the training images is the most challenging task. So Proposed system should have a effective mechanism for feature extraction.By creating a model using conventional neural network would be much more easier to implement the project.

References

[1] K. Parmar, et al., Facial-feature based Human-Computer Interface for disabled people,in Communication, Information Computing Technology (ICCICT), 2012 International Conference on, 2012

[2] Handbook of Face Recognition,S.Z Li and A.K.Jain,eds.Springer,2005

[3] Lei Zhang, Quanxue Gao and David Zhang, Block Independent Component Analysis for Face Recognition, 14th International Conference on Image Analysis and Processing, 2007.

[4] Vijay P.Shah, Nicolas H.Younan, Surya S.Durbha and Roger L.King, Feature Identification via a Combined ICA-Wavelet Method for image Information Mining, IEEE Geosciences and Remote Sensing Letters, vol.7, no. 1, January 2010.

[5] Abidin, T. and Perrizo, W. SMART-TV: A Fast and Scalable Nearest Neighbor Based Classifier for Data Mining. Proceedings of ACM SAC-06, Dijon, France, April 23-27, 2006. ACM Press, New York, NY, pp. 536-540.

[6] Nima Mesgarani, Shihab Shamma, and Malcolm Slaney, Speech discrimination based on multiscale spectrotemporal modulations, in IEEE International Conference on Acoustics, Speech and Signal Processing, Montreal, Canada, May 2004.

[7] Mark D. Skowronski and John G. Harris, Exploiting independent filter bandwidth of human factor cepstral coefficients in automatic speech recognition, Journal of Acoustical Society of America, vol. 116, no. 3, pp. 1774 - 1780, sept 2004.

[8] P. Viola and M. Jones, Rapid object detection using a boosted cascade of simple features,2001.

[9] Dan Duncan,Gautam Shine,Chris English.(2011)Facial Emotion Recognition in Real Time.International Journal of Computer Applications, vol. 15, pp. 37âĂŞ40, February 2011

[10] P. Abhang, S. Rao, B. W. Gawali, and P. Rokade, Article: Emotion recognition using speech and eeg signal a review, International Journal of Computer Applications, vol. 15, pp. 37âĂŞ40, February 2011.

[11] P. Ekman, Universals and cultural differences in facial ex- pressions of emotion. Nebraska, USA: Lincoln University of Nebraska Press, 1971.

[12] Dolly Reney,Dr.Neeta Tripaathi.(2015).An Efficient Method to Face and Emotion Detection.Fifth International Conference on Communication Systems and Network Technologies.

[13] E. M. Bouhabba,A. A. Shafie,R. Akmeliawati.(2011).Support vector machine for face emotion detection on real time basis Mechatronics (ICOM), 2011 4th International Conference.

[14] M. Turk and A. Pentland, Eigenfaces for recognition, Journal of Cognitive Neuro-science, pp. 110-118, 1991.

[15] G. Edwards, K. Cootes and C.Taylor, Face recognition using active appearance models,Lecture Notes in Computer Science, pp. 581-595, 1998.

[16] J. Goldstein, D. Harmon and B. Lesk, Identification of human faces, Proceedings of the IEEE, Vol.59, No 5, pp. 748âĂŞ760, May 1971.

[17] C. Corinna and V. Vapnik, Support-Vector Networks, Machine Learning, 1995.

[18] K. M. Rajesh M. Naveenkumar A robust method for face recognition and face emotion detection system using support vector machines International Conference on Electrical Electronics Communication Computer and Optimization Techniques (ICEECCOT) pp. 1-5 2016.

[19] Marius, Diedrick Pennathur, Sumita Rose, Klint. (2003). Face Detection Using Color Thresholding, and Eigenimage Template Matching.

[20] Hyung-Ji Lee , Wan-Su Lee, Jae-Ho Chung , Face recognition using Fisherface algorithm and elastic graph matching, IEEE International Conference on Image Processing ,Vol.1,pp: 998- 1001, October 2001.

[21] Raluca Boia,Radu Dogaru,Laura Florea. A comparison of several classifiers for eye detection on emotion expressing faces.Electrical and Electronics Engineering (ISEEE), 2013 4th International Symposium

[22] Carolann Cormier, MS, CCC-SLP, ATP. Establishing an Encoded Eye GazeCommunication system.

[23] R. A. King, T. C. Phipps, Shannon, Facial Emotion Recognition and approximation strategies, Computers Security,vol. 18, no. 5, pp. 445 453, 1999. V. Vapnick, The Nature of Statistical Learning Theory, SpringerVerlag, New-York, 1995.

Appendix A

Questions and Answers

Include atleast five questions and answers

Text of Appendix B is Here