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

Sainjal Poly

Reg No - 14004105

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 for the project titled **Real-Time Emotion Detection From Face** is a bonafide work carried out by **Sainjal Poly (Reg no-14004105)** 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

Apart from the efforts put in by us, the success of this project depends largely on the encouragement and guidelines of many others. We take this opportunity to express our gratitude to the people who have been instrumental in the successful completion of this stage of the project:

Mr.Paul Mundadan, chairman, FISAT Governing Body, who provided us with the vital facilities required by the project right from planning to completion of 1st Stage.

Dr. George Issac, Principal, FISAT for the amenities he provided, which helped us in the fulfillment of our project"s initial Stage.

Dr. Prasad J.C, HOD(CSE Dept), FISAT who always guided us and rendered his help in this phase of our project.

Mr.Pankaj Kumar G, for his constant encouragement and enthusiastic supervision and for guiding us with patience in 1st of the project .Without his help and inspiration, this would not have been materialized.

Mrs. Divya John ,Mrs Reshmi R ,Mrs. Preethi N P and Mr. Paul P Mathai and for their guidance and constant supervision as well as for providing necessary information regarding the project and also for their support in completing the 1st stage of the project The faculty of the CSE Dept., FISAT and Lab Instructors for providing us with the necessary Lab facilities and helping us throughout this project.

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

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Overview

The study of face and its features is an active research area from past few decades. Pose variation, illumination conditions, bad lighting etc., are still challenging factors faced by all algorithms. Face recognition and emotion detection system are the major applications of recognition system, in which many algorithms have tried to solve these problems. The face recognition is the basic part in modern authentication/identification applications; the accuracy of this system should be high for better results. Fisherface [1] 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. It is difficult for computer to find the face in light of the fact that the number of features in an image is extremely high. P.Viola and Jones [14] introduces different classifiers which groups the number of features into different classifiers. Using these classifiers, the different facial features are detected which can be used for further processing. But this method is limited to only frontal view of the face. Therefore in order to detect the facial features of the persons face under different poses, face landmark annotation [3] is required. For annotation of facial landmarks, the popular method active shape model (ASM) or open source software dlib can be used. Emotion detection is based on different expressions of face and these expressions are generated by variations in facial features.

Face emotion recognition [2] uses support vector machine [5] 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 [4] 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

Emotion recognition and affective intervention are nowadays well recognized desired features of Intelligent Tutoring Systems, with primary focus on such learner affective states as flow, boredom or frustration.Another areas of affective computing methods applications include testing driver stress , psychological diagnosis and training, neuro-biofeedback, emotion expression with avatars and more.The System evaluated three classifiers: Support Vector Machines, Modified Naive Bayes and Multilayer Perceptron. 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, used 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, applied a post-processing step on the output of the classifiers. With this approach, reached detection ratios of over 95% while keeping the rate of incorrect decisions below 5%. We conclude that 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 steps.

Chapter 2

Related Works

2.1 Face recognition using Fisherface algorithm and elastic graph matching

[1] Face recognition technique that effectively combines elastic graph matching (EGM) and the Fisherface algorithm. EGM as one of the dynamic link architectures uses not only face-shape 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 face recognition adopting 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[1]. 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 1990s, 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 procedure is impossible at real time

2.2 Overview Of Face Recognition System Challenges

Face recognition has become a valuable and routine forensic tool used by criminal investigators. Compared to automated face recognition, forensic face recognition is more demanding because it must be able to handle facial images captured under non-ideal conditions and it has high liability for following legal procedures. [6]The Author discusses recent developments in automated face recognition that impact the forensic face recognition community. Improvements in forensic face recognition through research in facial aging, facial marks, forensic sketch recognition, face recognition in video, near-infrared face recognition, and use of soft biometrics will be discussed. Finally, current limitations and future research directions for face recognition in forensics are suggested.

Face recognition is the ability to establish a subjects identity based on his facial characteristics. Automatic face recognition has been extensively studied over the past two decades due to its important role in a number of application domains, including access control, visual surveillance, and deduplication of government issued identity documents (e.g., passport and driver license), to name a few. Face recognition systems generally operate under one of two scenarios: verification or identification. In a verification scenario, the similarity between two face images is measured and a determination of either match or non-match is made. In an identification scenario, the similarity between a given face image (probe) and all the face images in a large database (gallery) is computed; the top (rank-1) match is returned as the hypothesized identity of the subject. Ideally, both of these scenarios are expected to operate in a lights out mode, i.e., the system makes an identity decision without requiring any human interaction.

2.3 One Millisecond Face Alignment with an Ensemble of Regression Tree

[8] Addresses the problem of Face Alignment for a single image. show how an ensemble of regression trees can be used to estimate the face's landmark positions directly from a sparse subset of pixel intensities, achieving super-realtime performance with high quality predictions. present a general framework based on gradient boosting for learning an ensemble of regression trees that optimizes the sum of square error loss and naturally handles missing or partially labelled data. show how using appropriate priors exploiting the structure of image data helps with efficient feature selection.[8] Different regularization strategies and its importance to combat overfitting are also investigated. In addition, analyse the effect of the quantity of training data on the accuracy of the predictions and explore the effect of data augmentation using synthesized data.

Present a new algorithm that performs face alignment in milliseconds and achieves accuracy superior or comparable to state-of-the-art methods on standard datasets. The speed gains over previous methods is a consequence of identifying the essential components of prior face alignment algorithms and then incorporating them in a streamlined formulation into a cascade of high capacity regression functions learnt via gradient boosting. We show, as others have , that face alignment can be solved with a cascade of regression functions. In our case each regression function in the cascade efficiently estimates the shape from an initial estimate and the intensities of a sparse set of pixels indexed relative to this initial estimate. Our work builds on the large amount of research over the last decade that has resulted in significant progress for face alignment. In particular, we incorporate into our learnt regression functions two key elements that are present in several of the succes

2.4 Computer vision for detection of body expressions of children

[11]This article is the result of an investigation to improve the communication with a case study that suffers Cerebral Palsy through the use of Computer Vision. At present, there is a 15% of the world's population that suffers some form of disability which prevents them from complying with the common activities of a normal person in a social environment. The technology can help to facilitate the implementation of processes that support people with special needs to improve their lifestyle; in this project the main objective is to improve the communication with the patient in order to facilitate the patient care. For conducting the investigation it was necessary the development of a prototype that detects body expressions using the OpenCV library with the Python programming language. The results are promising because the computer vision system is able to detect with high accuracy the following body patterns: headache 77%, happiness 75%, hunger 82%, fear 88% and recreation 77%. Finally, when a body pattern is detected it is communicated to the patient's caregiver through a mobile application.

Cerebral Palsy (CP) is the main cause of disability in children which prevents the development of movement and posture of the child who suffers it. The motor disorder of cerebral palsy is frequently accompanied in sensory, cognitive, communication and perceptual disorders of conduct, or by epilepsy. The CP is a common global problem so there are some families who have one member with special disabilities, having an incidence of 2 to 2.5 cases per 1,000 live births . Several of the children born under this condition do not communicate easily due to several factors, what comes to hinder direct understanding of the needs of the child; resulting in stress for both the family and the child with CP. Since there are several types of cerebral palsy (CP), it can affect in different ways on each person who possesses it; CP is not as a disease but as a set of syndromes, so it is advisable to investigate a specific case to get ideas and get the necessary conclusions for the present project.

2.5 Emotion Recognition Using Principal Component Analysis

[13]Emotion recognition plays vital role in Human Computer Interface. It focuses on facial expression to identify seven universal human emotions such as, happy, disgust, neutral, anger, sad, surprise and fear. This is carried out by trying to extract unique facial expression features among emotions using

Principal Component Analysis with Singular Value Decomposition and Euclidean Distance Classifier. Using public database Japanese Female Facial Expression (JAFFE) recognition is obtained nearly 78.57%. Recognition rate and Accuracy of various expressions using Principal Component Analysis alone and Principal Component Analysis with Singular Value Decomposition is compared.

Over the past decades, human-computer interaction together with computer vision has been an important field in computer study. The direct communication between the computer and human beings is matter of concerned. Much research has been conducted on improving and developing the interaction between human and the computer. One of the significant factors that contributed to increasing and developing the interaction between the computer and humans is studying the computers' ability to distinguish facial expressions for human. With emotion recognition systems, the computer will be able to assess the human expressions depending on their effective state in the same way that humans senses do. The intelligent computers will be able to judge, analyze and give response to person behaviors, expressions and moods.

The emotion recognition system applied in different areas of life such as security and surveillance, can predict the offender or criminal's behavior by analyzing the images of their faces that are captured by the control-camcorder. Furthermore, the emotion recognition system has been used in communication to make the answer machine more interactive with people. The answer machine has become more intelligent by analyzing the client's voice and dealing with the responses according to their emotions. Moreover, it is powerful in signed language recognition system that deals with the deaf and dumb people. The facial expression recognition system has a considerable impact on the game and entertainment field S.R.Khot Asso. Prof., Information Technology, D.Y. Patil College of Engineering Technology Kolhapur- 416006, India srkhot08@gmail.com besides its use to increase the efficiency of robots for specific military tasks, medical robots, and manufacturing servicing . Generally, the intelligent computer with facial expression recognition system has been used to improve our daily lives.

2.6 An Accurate Active Shape Model For Facial Feature Extraction

The active shape model (ASM) has been used successfully to extract the facial features of a face image under frontal view. However, its performance degrades when the face concerned is under perspective variations[15]. , a modified shape model is proposed to make the model represent a face more flexibly, under different orientations. The model of the eyes, nose and mouth, and the face contour are separated. An energy function is defined that links up these two representations of a human face. Three models are employed to represent the important facial features under different poses. A genetic algorithm (GA) is applied to search for the best representation of face images. Experiments show a better face representation under different perspective variations and facial expressions than the conventional ASM can.

Facial feature extraction is an important process in facial image analysis, which can be applied to face recognition and verification, animation, face image compression, clc. Important facial features include the face contour, eyes, nose. month. etc. Many approaches have been proposed for the extraction of these facial features. Snake s and deform able templates have been used to represent the face contour, eyes and mouth in a face image [16]. A global searching approach, active shape model (ASM) , which can adapt to any pride fined shapes more electrical and accurately, such as medical images, face images, hand gestures, etc. Modeling faces under different poses is a challenging problem, since the appearance of the racial features will significantly differ. ASM is constructed based on a linear combination of a set of 2D face appearances. which arc usually frontal view images. Consequently, if the input face is not of frontal view, the model cannot work properly. Therefore, in our approach, we model the face contour and tile facial features separately. For the facial features, three models are used to represent the features when the face is frontal view, turned left, and turned right, respectively. To c:o;tract the facial features in an image, we need to fit the defined face model to a face image. To search for the best match between the model and the face image optimally. the genetic algorithm (GA) is lIsed in our algorithm.

2.7 A Technique For The Measurement of Facial Movement

[18] A procedure has been developed for measuring visibly different facial movements. The Facial Action Code was derived from an analysis of the anatomical basis of facial movement. The method can be used to describe any facial movement (observed in photographs, motion picture film or videotape) in terms of anatomically based action units. The development of the method is explained, contrasting it to other methods of measuring facial behavior. An example of how facial behavior is measured is provided, and ideas about research applications are discussed. Although different in almost all other respects, most facial measurement techniques have shared a focus upon what is visible, what a rater can differentiate when he sees a facial movement[12]. An exception, Schwartz (in press) used electromyographic (EMG) measurement to study changes in muscle tone which are not visible. EMG could also be used to measure visible changes in muscle tone which do not involve a noticeable movement, but such work has not been done. EMG also could be employed to study visible movement, but we think it is unlikely that surface electrodes could distinguish the variety of visible movements which most other methods delineate. Vascular changes in the face are another aspect of facial behavior which can occur without visible movement and which, like muscle tonus, could be measured directly with sensors. No such work has been published on coloration or skin temperature although Schwartz

Research focused on the face has been impeded by the problems of devising an adequate technique for measuring the face. Over the years various procedures for facial measurement have been invented. Early work has rarely been cited by current investigators, e.g., FroisWittmann (1930), Fulcher (1942), Landis (1924), or Thompson (1941). More current approaches to facial measurement have varied in methodology, ranging from analogic notations of specific changes within a part of the face (Birdwhistell, 1970), to photographic depictions of movements within each of three facial areas (Ekman, Friesen, Tomkins, 1971), to verbal descriptions of facial gestalts (Young Decarie, Note 1). No consensus has emerged about how to measure facial behavior. No tool has been developed which has become the standard, used by all investigators. Each investigator has almost been in the position of inventing his own tool from scratch. The only exception has been that the category lists of facial behavior described by some human ethologists (Blurton-J ones, 1971; Grant, 1969; McGrew, 1972) have influenced other human ethologists studying children.

2.8 Face Detection using Color Segmentation Thresholding

[19]Color segmentation 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 ob- tained 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 on 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

2.9 Facial Emotion Recognition Based on Visual Information

[21]Facial emotion recognition (FER) 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 pro- vides 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 deep-learning 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 re- searches, 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.

2.10 Face Recognition and Face Emotion Detection System using Support Vector Machines

2.10.1 Introduction

[22] Pattern recognition aims at recognizing an object by its characteristic attributes. This chapter examines emotion recognition in the settings of pattern recognition problems. It begins with an overview of the well known pattern recognition techniques, and gradually demonstrates the scope of their applications in emotion recognition with special emphasis on feature extraction, feature reduction, and classification. The most common modality of emotion recognition is by facial expression analysis. Voice features used for emotion recognition include prosodic and spectral features. The chapter explores feature selection by single and multiple modalities and classification by neural, fuzzy, and statistical pattern recognition techniques. It also provides an overview of stimulus generation for arousal of emotion. Lastly, the chapter outlines the methods of performance analysis and validation issues in the context of emotion recognition.

2.10.2 Feature Vector

In order to obtain the features for training the classifiers, we explored the idea introduced in [4], where integral projections and TESPAR (Time Encoded Signal Processing And Recognition encoding are applied for eye detection. An integral projection (IP) is a simplified 1D view of a 2D image. The vertical IP is constructed by taking the sum of pixels on each row of the image (consider only grayscale images). The vertical projection will require sums over each column. Fig 2.1 contains a visual representation of the computation process. In addition to the two 1D signals presented, there are two extensions proposed by Feng in [5], namely the Variance Projection Functions. These functions are computed in a similar manner, along rows and columns and provide a measure of the variance of the pixel intensity distribution.

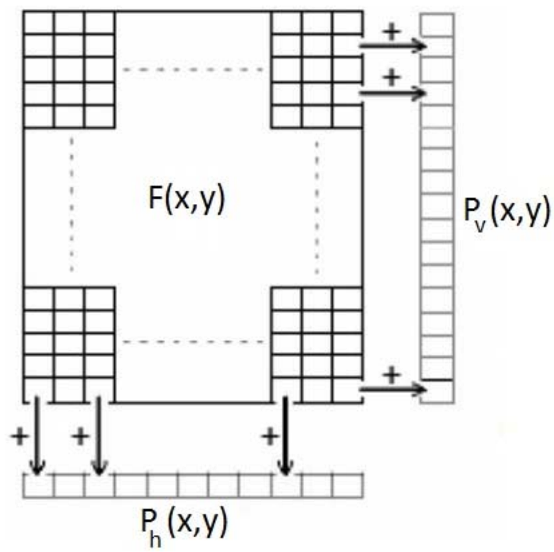

Figure 2.1. The computation of integral images

The four projections integrally projected signals are individually normalized in a symmetrical interval and then encoded by TESPAR method which was introduced by King et Al. [6] for speech signal processing. This technique will determine the zero crossings of the input signal, recording the length, the amplitude and the shape of the signal between two such intersections. By TESPAR encoding of each one of the four 1D integral projections and by concatenating the resulting signal values, the final set of 60 features for a given rectangular region from the image is obtained. The advantage of this encoding is that it preserves the important characteristics of the original 2D images, while reducing the complexity, thus allowing for a better generalization. These features will be computed for each image region and will be used for training and classification. aim in separating the feature vectors computed on eye regions from all the others computed on other parts of the face.

2.10.3 Neural Classifiers

The purpose of our work is to compare the performance of three classifiers and their ability to separate correctly the pixel regions which are close to the eye centers from those which are further away. The classifiers will be trained with features computed as previously described. The following methods are studied: Support Vector Machines, Naive Bayes and the Multilayer Perception.

2.10.3.1 Support Vector Machines (SVM)

The SVMs (introduced by Vapnik [7]) are considered to be among the top methods of supervised learning. They are generally used for binary classification, with the goal of finding a hyper planar surface which separates two data sets, by maximizing the margins between the two classes. While they are considered very powerful classifiers, the SVMs are sensible to outlets in the training set and to noise. The computational demands are also quite high for SVMs, both for processing power and memory usage. An example of SVM applied for emotion recognition in speech signals can be found in [8].

2.10.3.2 Modified Naive Bayes Classifier

The Bayes classifiers were introduced [9], [10] as a simplification of the Bayes theorem. In spite of the assumptions they make, they proved useful in many real world problems, especially when the data set is large and the training time must be fast (e mail filtering for example). Examples of uses can be found in [11], [12]. In our study, used a modified version of the Naive Bayes (NB) algorithm [13], in which changes were performed to make it more suitable for embedded systems. The modifications target the reduction of computational complexity [14] firstly, a tuning parameter was added which works similarly to the radius for networks with radial basis functions; secondly, the exponential function was approximated by a piecewise linear function. The new implementation of the NB classifier has a similar accuracy to the standard one, while the speed is greatly improved.

2.10.3.3 Multilayer Perceptron (MLP)

The perceptron is the simplest type of neural network, its goal being to classify linearly separable shapes [15]. Multiple perceptrons connected in parallel and then in a series will create a Multilayer Perceptron network [16], [17]. Due to the inner layers, this kind of networks grants good generalization, allowing them to handle complex problems. While the training of the MLP can become computationally intensive for large volumes of data, the execution time for classification is very short, making them compatible with real time applications.

2.10.4 Face Recognition

Face recognition is a process of identification of a persons face in an image or video which includes several steps need to be processed. Figure 2.2 shows the block diagram of face recognition system, which includes face detection, face extraction and face matching. Recognition is based on the stored image data of the different group of persons. Input images are of any type can be used for recognition

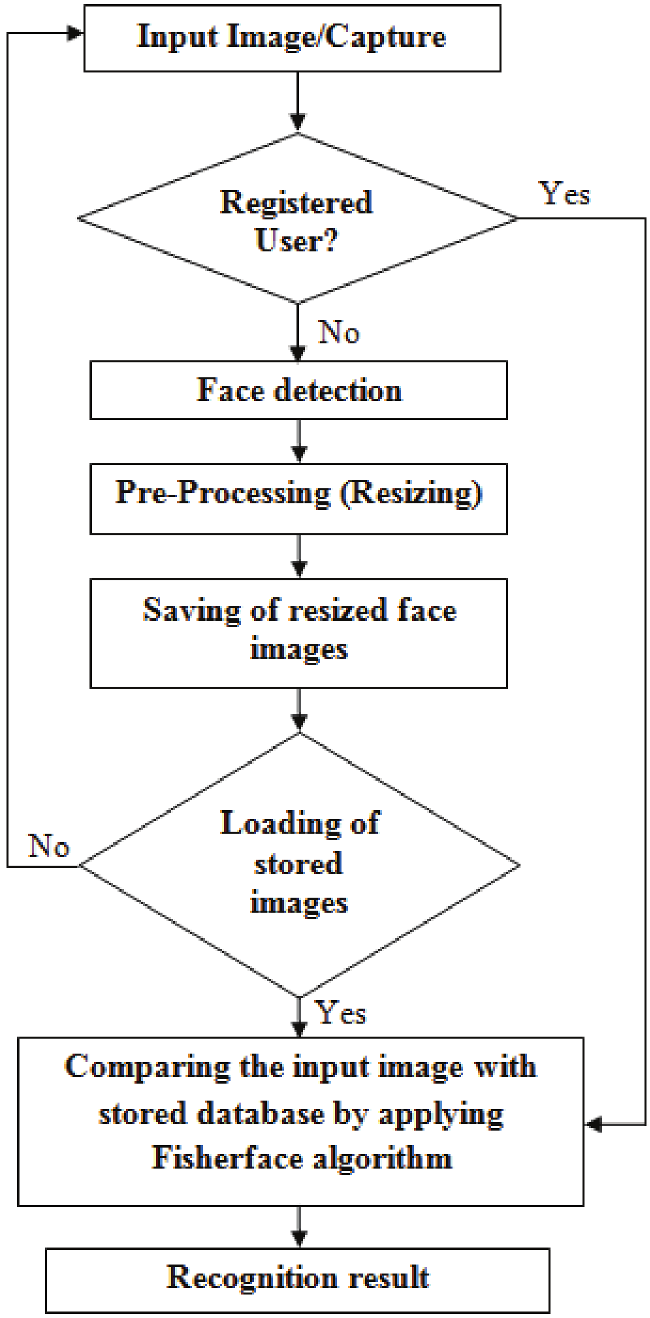

Figure 2.2. Real time Face recognition system

- Still images
- Video frames or video stills
- Video

Input image is subjected for face detection to detect the face. Detected face is then extracted from the image and these images are saved as a database. Saved images are used to compare with the input image. The matching of input image is performed to identify the users identity. The recognition result gives identification of the person (particularly his/her name). Figure 2.2 depicts the step by step architecture developed for face recognition

The algorithm is designed in such a way that, if the person is recognizing for the first time then the system considers him as a new user and performs each step of operation. But if the person/user data is already stored then it is considered as Registered user and it performs only matching operation to recognize the user identity. OpenCV contains cascade classifiers in which Viola Jones face detection algorithm is implemented. By using these classifiers the face region is detected from the image by fig 2.3. It classifies the images into positive and negative images respectively.

The images consists of face region is considered as positive and the images without face as negative images. These negative images are ignored for further processing. The stored images consists of face images of dimension 273x273, the more number of the images higher the recognition rate. If the stored images consists false images or wrong extensions then recognition is not possible. Therefore care should be taken while capturing input images. Fisherface algorithm is applied for classification of different users. Fisherface algorithm generates the fisherfaces of each image that are used for recognition. In Fisherface algorithm, it performs leave one out cross validation to validate the user identification.

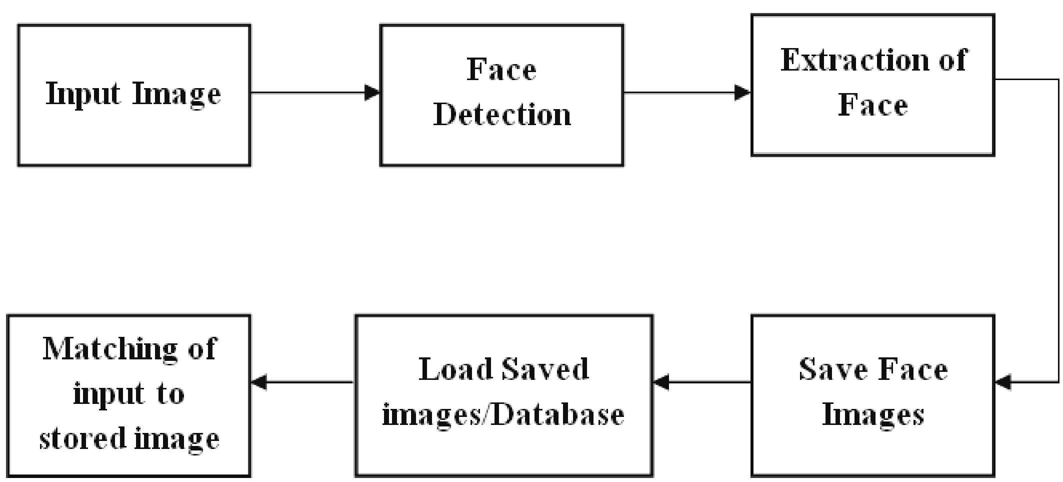

Figure 2.3. Block diagram of Face recognition system.

2.10.4.1 Voila-Jones Face Detection

Voila 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

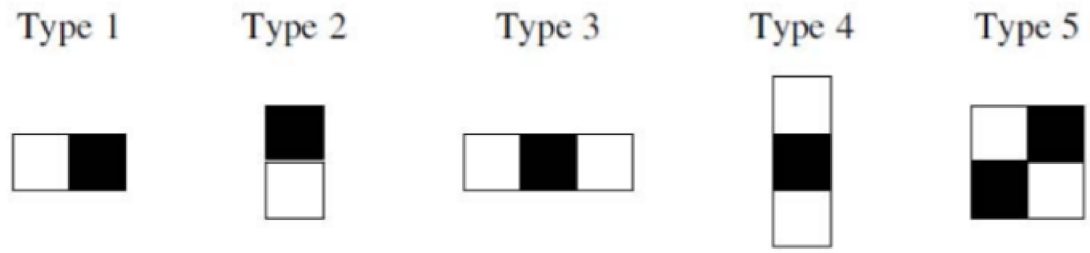

Figure 2.4. The different type of features

out to be rather time consuming due to the calculation of the different size images. Contrary to the standard approach Viola Jones 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 [6] 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 in constant time. The Viola 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 2.4

2.10.5 Emotion Detection

Face emotion detection is used to predict the emotion state of the person based on their face expressions. Here input images are classified into two types,

- Training images
- Testing images.

Training images are used for training of classifier. Testing images are used to verify the algorithm by predicting the different emotions of the face. Expression analysis is the major part of the emotion detection, the schematic of expression analysis for classifying different emotions

PCA is applied to training images to reduce the dimensionality. Because training images are more compared to testing and if the dimension is high then the time taken for processing will be more. Support vector machine classification is done for classifying different emotions namely, Happy, Sad, Angry, Fear, Disgust and Surprise. The emotion detection system detailed flow diagram is shown in figure 2.5

Facial features such as eyes, nose, lips and face contour are considered as the action units of face and are responsible for creation of expressions on face, are extracted using open source software called dlib. SVM classifier compares the features of training data and testing data to predict any emotion of the face.

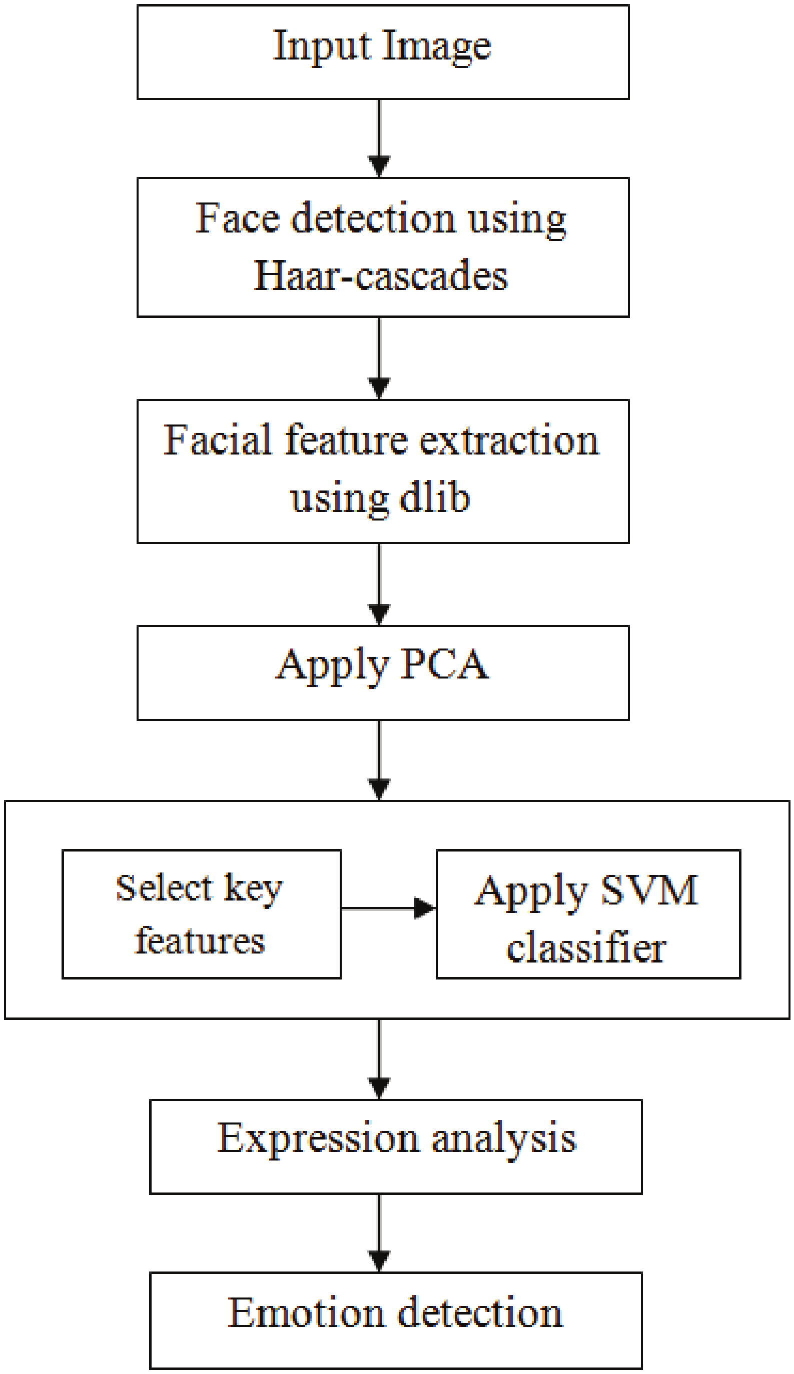

Figure 2.5. Workflow of Emotion detection system

Here facial features are considered as the key points which are used for training and testing. Support vector machine is the supervised learning method of machine learning. Machine learning algorithms are advantageous over other algorithms, because of less error rate and faster results. LinearSVC which is also called as MultiSVM [11] is used for classification

2.10.6 Testing SVM

For SVM, computed the TESPAR features on the training database, from which selected 80% of the positive examples and 80% of the negatives. After a first training of the classifier, the set of negatives was expanded with some of the negatives incorrectly classified as positives. The second round of training used the same set of positives, but more negatives this two stage training was found to yield better results than a single training. used a non linear SVM classifier with a radial basis function, implemented in C++ and accessed through the Matlab wrappers [21]. performed multiple tests by changing the gamma parameter of the network, the choice of this parameter having a significant impact over the learning ability of the system.

Table 2.1. Time Estimation

Type	**Time Taken(SFC)**
Face detection	0.0844
Facial feature extraction	0.9216
Classification using SVM	0.1956
Emotion Detection	0.1994

A first conclusion was that the best results were obtained for the percentage P=80% (this overlap percentage defines a positive feature). The results with P=95% were the worst, making it clear that the network cannot learn due to insufficient number of positive examples. Regarding the database content, for SVM the best results were obtained for trainings done on the mixed database, which contained both neutral and emotion images, proving that the features from the neutral images help improve the detection on the images expressing emotions On the image database, the measured accuracy is up to 98% if the T eye threshold is 0.1 (we consider correct to have the detected eye center to be in the region of the iris) and 86.5% if T eye threshold is 0.05. (the detected eye is in the pupil region). The testing also showed that the use of adaptive post processing will improve the performance by a few percent (2 3% in our experiments). conclude that this post filtering of the results influences the system in such manner that the best accuracies are not necessarily obtained using the classifiers that provided the highest performance.

2.10.7 Experimental Results of Face Recognition

The system is tested on datasets which consists of wide range of face images with different expressions, poses, illumination conditions and genders. used CK and CK+ database [10] for training of emotion detection system. Algorithm is tested on IMM database [7] and also our own test images. Both databases are open source and our algorithm performed well on both datasets.

Figure 2.6. Real-Time Face recognition results.

For face recognition, have used a webcam for capturing of faces. The implemented algorithm is capable of recognizing different persons in a single window as shown in figure 2.6. If the recognition environment is under proper lighting condition and less background noises then the recognition rate will be high.

2.10.8 Experimental Results of Emotion Detection

The images of dimension 640x480 are used for testing of emotion detection system. For training, the images from Ck and Ck+ database are used. 320 images of different expressions are used for training. For testing of the algorithm, used the images which are captured using webcam. 50 to 60 different set of images of different persons are used in testing. MultiSVM classifier is used for classification of different emotions. One All SVM classifier is used for training of different classes of expressions. Dlib is used to extract the facial features, the experimental results shows the detection of different emotions. PCA is used to reduce the dimensionality; it finds the small number of eigenfaces. These eigenfaces should span a space that is required to represent a face. Figure 5.2 shows the detected results of different emotions.

The emotions can be classified as positive and negative, these can be used to understand the mental condition of the person.

The implementation is done using OpenCV and python along with additional dependencies like dlib, scikit learn and scimage. Table 2.1 shows the time estimation of different detections performed. The time taken for each process is obtained using the time function of the python.

Face emotion recognition [2] uses support vector machine [5] 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 [4] 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.

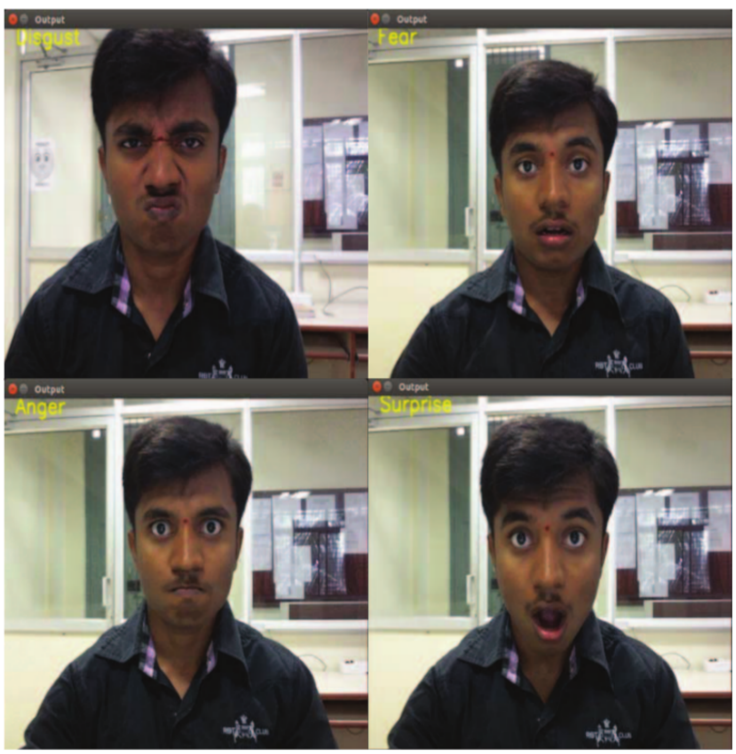

Figure 2.7. Detection of Disgust, Fear, Anger and Surprise faces

2.11 Real-time Convolutional Neural Networks for Emotion and Gender Classification

2.11.1 Introduction

[23] The success of service robotics decisively depends on a smooth robot to user interaction. Thus, a robot should be able to extract information just from the face of its user, e.g. identify the emotional state or deduce gender. Interpret ing correctly any of these elements using machine learning (ML) techniques has proven to be complicated due the high variability of the samples within each task [1]. This leads to models with millions of parameters trained under thousands of samples [2]. Furthermore, the human accuracy for classifying an image of a face in one of 7 different emotions is 65% + 5% [3]. One can observe the difficulty of this task by trying to manually classify the FER-2013 dataset images in within the following classes {"angry", "disgust", "fear", "happy", "sad", "surprise", "neutral"}. In spite of these difficulties, robot platforms oriented to attend and solve household tasks require facial expressions systems that are robust and computationally efficient. More- over, the state-of-the-art methods in image-related tasks such as image classification [4] and object detection are all based on Convolutional Neural Networks (CNNs). These tasks require CNN architectures with millions of parameters; therefore, their deployment in robot platforms and real-time systems. becomes unfeasible. The proposed system implement a general CNN building framework for designing real-time CNNs. The implementations have been validated in a real-time facial expression system that provides face-detection, gender classification and that achieves human-level performance when classifying emotions. This system has been deployed in a care-O-bot 3 robot, and has been extended for general robotplatforms and the RoboCup@Home competition challenges. Furthermore, CNNs are used as black-boxes and often their learned features remain hidden, making it complicated to establish a balance between their classification accuracy and unnecessary parameters. Therefore, implemented a real- time visualization of the guided-gradient back-propagation proposed by Springenberg [11] in order to validate the features learned by the CNN.

2.11.2 Design of Modules

2.11.2.1 Mini-Xception model

The second model is inspired by the Xception architecture. This architecture combines the use of residual modules and depth-wise separable convolutions as shwon in figure 2.8.Residual modules modify the desired mapping between two subsequent layers, so that the learned features become the difference of the original feature map and the desired features. Consequently, the desired features H(x) are modified in order to solve an easier learning problem F (X) as shown in eqn 2.1

$$H(x) = F(x) + x \tag{2.1}$$

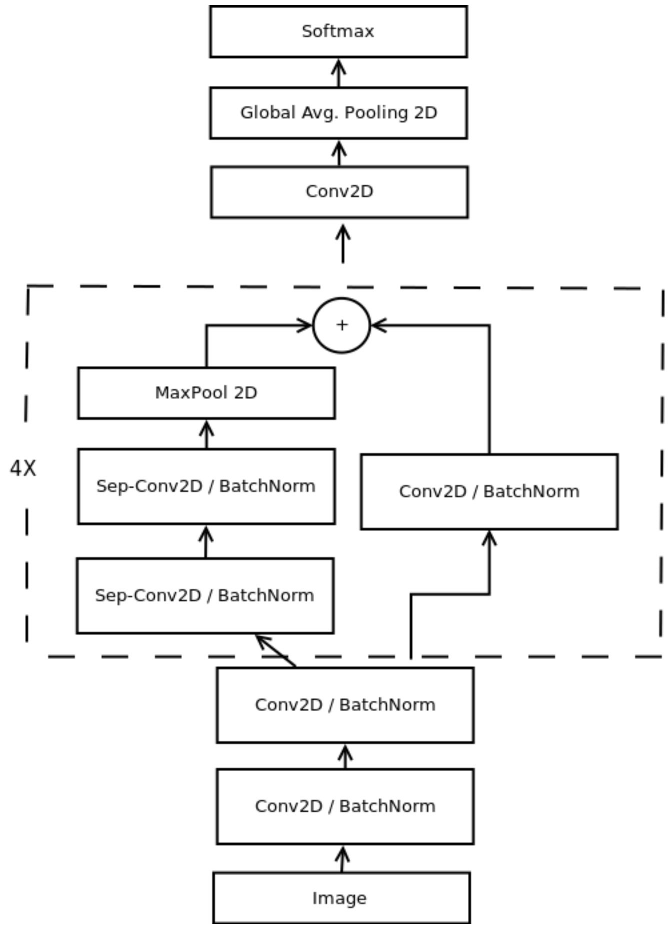

Figure 2.8. Proposed model for real-time classification

Since our initial proposed architecture deleted the last fully connected layer, we reduced further the amount of parameters by eliminating them now from the convolutional layers. This was done trough the use of depth-wise separable con volutions. Depth-wise separable convolutions are composed of two different layers: depth-wise convolutions and point- wise convolutions as shown in figure 2.9. The main purpose of these layers is to separate the spatial cross-correlations from the channel cross- correlations[1].Depth-wise separable convolutions reduces the computation with respect to the standard convolutions by a factor of $\frac{1}{N} + \frac{1}{D^2}$. A visualization of the difference between a normal Convolution layer and a depth-wise separable convolution can be observed Our final architecture is a fully-convolutional neural network that contains 4 residual depth-wise separable convolutions where each convolution is followed by a batch normalization operation and a ReLU activation function. The last layer applies a global average pooling and a soft-max activation function to produce a prediction. This architecture has approximately 60, 000 parameters; which corresponds to a reduction of 10X wen compared to our initial naive implementation, and 80X when compared to the original CNN. Displays our complete final architecture which we refer to as mini-Xception.

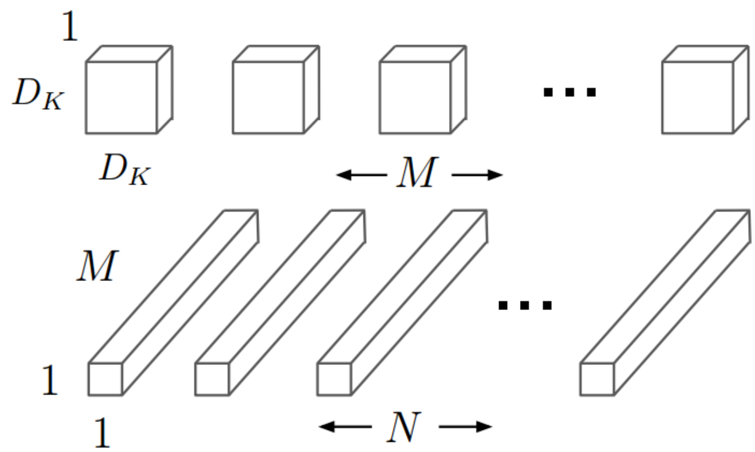

Figure 2.9. Standard convolutions vs Depth-wise separable convolutions

Here system tested this architecture in the FER-2013 dataset and we obtained the same accuracy of 66% for the emotion classification task. Our final architecture weights can be stored in an 855 kilobytes file. By reducing our architectures computational cost it is able to join both models and use them consecutively in the same image without any serious time reduction. Our complete pipeline including the openCV face detection module, the gender classification and the emotion classification takes 0.22 + 0.0003 ms on a i5-4210M CPU. This corresponds to a speedup of 1.5X when compared to the original architecture of Tang. We also added to our implementation a real-time guided back-propagation visualization to observe which pixels in the image activate an element of a higher-level feature map. Given a CNN with only ReLUs as activation functions for the intermediate layers, guided-back propagation takes the derivative of every element $(x, y)$of the input image I with respect to an element (i, j) of the feature map f L in layer L. The reconstructed image R filters all the negative gradients; consequently, the remaining gradients are chosen such that they only increase the value of the chosen element of the feature map. .

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 other's 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 this paper, we propose an emotion detection system with facial features using a Bayesian network. 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, we propose an emotion detection system 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, player's affect state can change even radically from session to session making almost impossible to predict the current user emotions at the development stage.

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.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 person, 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.

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. Our approach is outlined next. 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.

Web development is already on the path towards more personalization. As emotion recognition technology becomes more sophisticated and more deeply embedded in our array of devices, it will become expected that our computers and phones provide us with a continual progression of customized triggers and messaging. The technology will be found even in future car dashboards, refrigerator doors, and conference room walls essentially any surface will become a possible means for detection of emotions.

Social media will constantly focus on each user's emotions. For example, in the future expect Facebook's algorithm to focus just as much on one's emotional reactions as it does to one's historical click behavior, providing a unique social environment that goes far beyond prediction of the types of posts, pages, and ads one would like. Expect Facebook Ads to provide advertisers with the ability to hyper target not only based on age, geography, and job titles, but also on the individual's emotional state or progression of emotional states.

Online marketing will likely evolve into sequential experiences, with deeper engagement upon recognition of positive emotional reactions. You can also expect more deeply embedded forms of marketing similar to product placement. Ultimately, expect emotion recognition to be just another core component of marketing, similar to how digital marketing is now really just marketing.Emotion recognition technology is clearly bringing about a revolution in marketing.

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 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 it would be much more easier to implement the project.

References

[1] 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

[2] Adrian Dinculescu,Cristian Vizitiu,Alexandru Nistorescu,Mihaela Marin,Andreea Vizitiu, *Novel Approach to Face Expression Analysis in Determining EmotionalValence and Intensity with Benefit for Human Space Flight Studies* 5th IEEE International Conference on E-Health and Bioengineering - EHB, pp:1 - 4, November 2015

[3] Rajesh K M, Naveenkumar M, *An Adaptive-Profile Modified Active Shape Model for Automatic Landmark Annotation Using Open CV*, International Journal of Engineering Research in Electronic and Communication Engineering (IJERECE), Vol.3, Issue.5, pp:18-21, May 2016.

[4] Samiksha Agrawal,Pallavi Khatri, *Facial Expression Detection Techniques: Based on Viola and Jones algorithm and Principal Component Analysis*, Fifth International Conference on Advanced Computing Communication Technologies, pp:108-112, February 2015.

[5] Brahim A. Adeyanju,Elijah O. Omidiora,Omobolaji F. Oyedokun, *Performance Evaluation of Different Support Vector Machine Kernels for Face Emotion Recognition*, SAI Intelligent Systems Conference,pp: 804 806, November 2015.

[6] Ambika Ramchandra , Ravindra Kumar, *Overview Of Face Recognition System Challenges* , International journal of scientific technology research , Vol.2,pp: 234-236, August 2013.

[7] M. M. Nordstrom,M. Larsen,J. Sierakowski,M. B. Stegmann, *The imm face database-an annotated dataset of 240 face images , Elsevier, DTU Informatics, Tech. Rep. , 20*04.

[8] V. Kazemi,J. Sullivan, *One Millisecond Face Alignment with an Ensemble of Regression Trees*, IEEE Conference on Computer Vision and Pattern Recognition,pp:1867 1874, 2014.

[9] Abu Sayeed Md. Sohail,Prabir Bhattacharya, *Classifying Facial Expressions Using Point-Based Analytic Face Model and Support Vector Machines* , IEEE International Conference on Systems, Man and Cybernetics,pp:1008-1013, October 2007.

[10] Patrick Lucey, Jerey F. Cohn , Takeo Kanade, *The Extended Cohn- Kanade Dataset (CK+): A complete dataset for action unit and emotion specified expression* , IEEE Computer Society Conference on Computer Vision and Pattern Recognition, pp: 94 to 101,June 2010.

[11] Jan Erik Solem, *Programming Computer Vision with Python*, First Edition , ISBN 13, July 2012.

[12] Ketki R. Kulkarni,Sahebrao B. Bagal, *Facial Expression Recognition* ,2015 Annual IEEE India Conference (INDICON), pp:1 to 5, 2015.

[13] Ajit P. Gosavi,S.R.Khot, *Emotion Recognition Using Principal Component Analysis with Singular Value Decomposition* , International Conference on Electronics and Communication System (lCECS - 2014),pp:1 to 5, , February 2014.

[14] P. Viola,M. Jones, Rapid object detection using a boosted cascade of simple features, Proceedings of the 2001 IEEE Computer Society Conference ,Vol.1,pp: 511 to 518, 2001.

[15] K. W. Wan,K. M. Lam,K. C. Ng, *An accurate active shape model for facial feature extraction, Pattern recognition letters* , Vol.26,pp:2409 to 2423, 2005

[16] Dibeklioglu, H. and Salah, A. A. and Gevers, T. *Are You Really Smiling at Me? Spontaneous Versus Posed Enjoyment Smiles* , European Conference on Computer Vision (ECCV 2012), vol. 7574, pp 525 to 538, 2012.

[17] Deriso, D, Susskind, J, Krieger, L, Bartlett MS (2012). *Emotion Mirror: A novel intervention for autism based on real-time expression recognition* . European Conference on Computer Vision (ECCV 2012), Vol 7585, pp. 671 to 674.

[18] P. Ekman, W. Friesen, Facial Action Coding System: *A technique for the measurement of facial movement* , Palo Alto, U.S. CA, Consulting Psychologists Press, 1978.

[19] Laura Florea, Corneliu Florea, Ruxandra Vranceanu, Constantin Vertan, *Zero-Crossing Based Image Projections Encoding for Eye Localization*, 20th European Signal Processing Conference (EUSIPCO), 2012.

[20] G. C. Feng, P. C. Yuen, *Variance projection function and its application to eye detection for human face recognition*, Pattern Recognition Letters, vol. 19, no. 9, pp. 899 to 906, July 1998.

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

[22] Rajesh K M,*A Robust Method for Face Recognition and Face Emotion Detection System using Support Vector* 978-1-5090-4697-3/16 31.00 2016 IEEE Machines

[23] Octavio Arriaga Hochschule Bonn-Rhein-Sieg Sankt Augustin Germany *Real-time Convolutional Neural Networks for Emotion and Gender Classification* IEEE International Conference on Systems,Man and Cybernetics,pp:1008-1013, October 2007.