

FACIAL EXPRESSION RECOGNITION VIA DEEP LEARNING

*Submitted in partial fulfillment of the requirements for the award of
Bachelor of Engineering degree
in Information Technology*

By

FEROSE MOOSHA F (Reg No - 39120031)

ARUNACHALAM R (Reg No - 39120014)



DEPARTMENT OF INFORMATION TECHNOLOGY

SCHOOL OF COMPUTING

SATHYABAMA

INSTITUTE OF SCIENCE AND TECHNOLOGY

(DEEMED TO BE UNIVERSITY)

Accredited with Grade "A" by NAAC | 12B Status by UGC | Approved by AICTE

JEPPIAAR NAGAR, RAJIV GANDHISALAI,

CHENNAI - 600119

MAY - 2023



SATHYABAMA

INSTITUTE OF SCIENCE AND TECHNOLOGY

(DEEMED TO BE UNIVERSITY)

Accredited with —A+ grade by NAAC

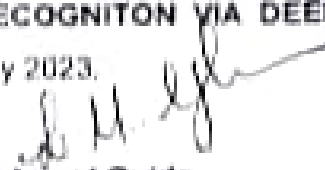
Jeppiaar Nagar, Rajiv Gandhi Salai, Chennai - 600 119

www.sathyabama.ac.in

DEPARTMENT OF INFORMATION TECHNOLOGY

BONAFIDE CERTIFICATE

This is to certify that this Project Report is the bonafide work of **FEROSE MOOSHA F (39120031)** and **ARUNACHALAM R (39120014)** who carried out the Project Phase-2 entitled "**FACIAL EXPRESSION RECOGNITION VIA DEEP LEARNING**" under my supervision from January 2023 to May 2023.


Internal Guide

Dr. MARY GLADENCE, M.E., Ph.D.


Head of the Department

Dr. R. SUBHASHINI, M.E., Ph.D.

Submitted for Viva voce Examination held on 10.05.2023


Internal Examiner


External Examiner

DECLARATION

I, **FEROSE MOOSHIA F** hereby declare that the Project Phase 2 Report entitled **FACIAL EXPRESSION RECOGNITION VIA DEEP LEARNING** done by me under the guidance of Dr. **MARY GLADENCE ,M.E.,Ph. D** is submitted in partial fulfillment of the requirements for the award of Bachelor of Technology degree in Information Technology

DATE: 10.05.2023

PLACE: Chennai


SIGNATURE OF THE CANDIDATE

ACKNOWLEDGEMENT

I am pleased to acknowledge my sincere thanks to **Board of Management of SATHYABAMA** for their kind encouragement in doing this project and for completing it successfully. I am grateful to them.

I convey my thanks to **Dr. T. Sasikala M.E., Ph.D.** Dean, School of Computing, **Dr. R. Subhashini, M.E., Ph.D.** Head of the Department of Information Technology for providing me necessary support and details at the right time during the progressive reviews.

I would like to express my sincere and deep sense of gratitude to my Project Guide **Dr. MARY GLADENCE M.E., Ph.D.** for her valuable guidance, suggestions, and constant encouragement paved way for the successful completion of my phase-1 project work.

I wish to express my thanks to all Teaching and Non-teaching staff members of the **Department of Information Technology** who were helpful in many ways for the completion of the project.

ABSTRACT

The human face has peculiar and specific characteristics, therefore it becomes difficult in understanding and identifying the facial expressions. It is easy to identify the facial expression of person in any image sequence. If we look to automated recognition system, however, the systems available are quite inadequate and incapable of accurately identify emotions. The area of facial expression identification has many important applications. It is an interactive tool between humans and computers. The user, without using the hand can go-ahead with the facial expressions. Presently, the research on facial expression is on the factors i.e., sad, happy, disgust, surprise, fear and angry. This paper aims to detect faces from any given image, extract facial features (eyes and lips) and classify them into 6 emotions (happy, fear, anger, disgust, neutral, sadness). The training data is passed through a series of filters and processes and is eventually characterized through a Support Vector Machine (SVM), refined using Grid Search. The testing data then tests the data and their labels and gives the accuracy of classification of the testing data in a classification report. Various approaches, including passing the training images through Gabor filter, or transforming images using Histogram of Oriented Gradients (HOG) and Discrete Wavelet Transform (DWT) for better classification of data are implemented. The best result achieved so far is by passing the training images through Histogram of Oriented Gradients (HOG), followed by characterization by SVM, which gives an average precision of 85%. Moreover, we propose identifying epochs, which are the instants at which excitation is maximum, during the emotion to improve the system's accuracy. We used the zero-time windowing method to extract instantaneous spectral information using the numerator group-delay function to accurately detect the epochs in each emotional state. Different classification scheme was defined using QDC and RNN and evaluated using the DEAP database. The experimental results showed that the proposed method is highly competitive compared with existing studies of multi-class emotion recognition. The average accuracy rate exceeded 89%. Compared with existing algorithms dealing with 9 emotions, the proposed method enhanced the accuracy rate by 8%. Moreover, experiment shows that the proposed system outperforms similar approaches discriminating between 3 and 4 emotions only. We also found that the proposed method works well, even when applying conventional classification algorithms.

TABLE OF CONTENT

CHAPTER NO	TITLE		Page No
	ABSTRACT		v
	LIST OF FIGURES		vii
1	INTRODUCTION		8
2	LITERATURE SURVEY		11
	2.1	Inferences from Literature Survey	11
	2.2	Open problems in Existing System	15
3	EMOTION DETECTION PROCESS AND APPLICATIONS		17
	3.1	Modules in FERDL	17
	3.2	Framework Overview	18
	3.3	Software Requirements in FEDDL	19
	3.4	Key frame extraction from input video	20
	3.5	Use cases of Emotion Detection	20
	3.6	Methods and Approaches	21
	3.7	6 Types of Datasets	22
	3.8	Machine learning's Application	23
4	DESCRIPTION OF PROPOSED SYSTEM		26
	4.1	Selected Methodology or process model	26
	4.1.1	Process Model for Emotion Detection	27
	4.1.2	Facial Recognition with Music Recommendation	34
	4.2	Architecture / Overall Design of Proposed System	35
	4.3	Emotion recognition by CNN	35
	4.4	Convolution Filter	37
5	EXPERIMENTAL ANALYSIS AND EVALUATION		38
	5.1	Experimental Environment	38
	5.2	Structure of CNN	39
	5.3	Recognizing Emotion from Live Videos	39
	5.4	Implementation	41
	5.5	Performance Analysis and Evaluation	43
	RESULT AND CONCLUSION		46

	REFERENCES	48
--	------------	----

LIST OF FIGURES

FIG NO	FIGURE NAMES	PAGE NO
3.2.1	IMAGE RECOGNITION FRAMEWORK	18
4.2.1	ARCHITECTURE OF FERDL	35
5.3.1	IMAGE FEATURE EXTRACTION MODEL	39
5.3.2	PROCESS MODEL OF FER	40

CHAPTER 1

INTRODUCTION

In the field of education, existing virtual learning environments successfully simulate interaction at a cognitive level during traditional teaching processes. In the process of human-computer interaction, teachers naturally expect a facial recognition system to have the ability to detect, analyze and process emotions in order to get good teaching effect, such as perception, understanding and expressing emotions. If in virtual learning environments, students show different expressions in videos, teachers can identify whether the students understand their teaching content according to students' different expressions, and can adjust their teaching programs. In virtual environments, students' emotional data can be obtained from a series of video frames. Therefore, to achieve real time understanding of students' emotions, accuracy and time needs to be balanced. To improve accuracy, more video frames are required, increasing computation time. On the other hand, in pursuit of efficiency, we would reduce the accuracy which means collection less feature data. Experimental results from current solutions show that in some cases accuracy is higher, but time spent is also high, such as Salman et al who used Decision Trees to identify Facial Expressions. Other solutions have different limitations; for example, Chu et al [4] provided low-cost facial expressions on a mobile platform by just extracting lip features, spending less time but achieving less accuracy. To find the best solution for emotion recognition in learning virtual environment, both accuracy and efficiency needs to be achieved. Optimization of recognition of emotional changes in the facial features of online learners, will allow teachers to adjust teaching strategies and methods, give real-time feedback to students, and achieve the best teaching quality. The goal is to find the best solution for emotion recognition based on facial recognition in virtual learning environments, in real time. To achieve this goal, it is necessary to improve the accuracy and efficiency of facial recognition systems. Face is the most visible part of the human body, with which the identity of the person as well as their age or even their gender can be also be detected. As many studies in the face recognition conveys, facial emotion recognition rises and become well-known. Scientific research showed a very high potential for facial emotion recognition

to be applied in a vast range of applications. Many companies offering a vast range of consumer products take advantage of face emotion recognition technologies for recognizing customer feedback, identifying potential customers. These facial expressions become more valuable to the field of machine learning. In the field of education, face emotions are used to determine the level of understanding of the students. Teachers have the freedom to teach their students in a manner that they are comfortable. They deliver the lessons using different platforms, digital or traditional way. The main objective is that at the end of each lesson, the students should understand the lesson. Using face emotion recognition, teachers can evaluate whether the students followed the lecture or not. Based on the feedback determined using the face emotion patterns, teacher may change his/her strategy in delivering the lecture. Hospitals and healthcare institutions may also rely on the development of face emotion recognition. As patients, become immobilized, their ability to reach out for medical attention become difficult, and also the evaluation of the patients' pain condition relies on continuous monitoring of a medical staff. With face emotion recognition, what the patients are feeling can be detected and evaluated. According to the studies, there are 6 basic emotions that can be read by the current state of machine learning models namely – anger, sadness, happiness, disgust, fear surprise and neutral. Removing the neutral from the list, the 5 remaining emotions can be grouped into positive (surprise and happiness) and negative (anger, sadness, disgust, and fear). The formation of these basic emotions can still be enhanced to reveal deeper and complex emotions. There are several commercially available face recognition technologies including Affectiva Affectix, Microsoft Cognitive Services module Emotion Recognition, Amazon Recognition Face Analysis, and Neurodata Lab Emotion Recognition attained their identities in the field of facial emotion recognition [2]. Despite of these, many independent researchers continue to devise their models to address unseen drawbacks. Emotion is one of the very few words in the English language that do not have a concrete definition and it is understandable. It is abstract. Yet, almost every decision we have ever made in our lives is driven by emotion. Marketing research has proven that predicting sentiments correctly can be a huge source of growth for businesses and that is what we will be working on today — Reading Emotions. In the world of data and machine learning, this concept falls under the umbrella of cognitive systems. Let us try to decode the science behind Emotion Recognition Algorithms, and build one for ourselves.

What exactly is a cognitive emotion detection algorithm trying to accomplish? The idea is to replicate the human thought process based on training data (in the form of images and videos of humans) and try to segment the emotions present in this data. To perform our analysis in this chapter we will be concentrating on pre-recorded images and videos that showcase an emotion, but the same can also be implemented on a live stream of the video feed for real-time analytics. Several convoluted use cases justify the use of computer vision in today's industry. Some very basic uses are, in an examination conducted online. A web camera can read the expressions of the user to interpret their state of mind. This is also useful in testing emotional strength for pilots and race car drivers before they move to the cockpit for the final drive. Many robots today, including voice assistants like Alexa and Siri, are successfully able to mimic human behavior and speak back sentimentally. This also traces to cognitive therapy, which deals with stress and anxiety disorders among war veterans and equity traders, who are constantly under emotional strain. Sentiment Analysis and Face Detection, individually have numerous use-cases in today's world. We see object detection algorithms in public parking lots, traffic monitoring systems, etc. that take images of people driving vehicles to keep records. Sentiment Analysis is furthermore used in therapy where physical meetings of the therapist and their patient are not possible. The study of human cognition has also evolved medicines. On the technological front, virtual assistants, profile evaluation assistants, and automation bots are built to mimic the actions of humans and replace them with the hope of increasing accuracy and decreasing errors. It is therefore a very important part of the Artificial Intelligence inspired world we live in today. A more engrossing and complicated approach to computer vision is by using cloud-based algorithms like Azure Cognitive Services or Deep Learning mechanisms, which we have not covered in this story, but could come in handy for complex scenarios. Each frame analyzed by this function is stored as a separate image by the algorithm in the root directory folder where the code is running. Also, this function later creates a replica of the original video by placing a box around the face and showing live emotions within the video. So essentially, the working of any algorithm for both videos and images is the same. The only additional step in processing videos is splitting the video into all of its individual frames and then applying the image processing algorithms to it.

CHAPTER 2

LITERATURE SURVEY

2.1 INFERENCES FROM LITREATURE SURVEY

This problem statement has been extensively studied over the past 5 years by researchers and automotive companies in a bid to create a solution, and all their solutions vary from analyzing various patterns of distractive habits to analyzing health vitals of the driver.

**1)Title: Capturing-the-Invisible (CTI): Behavior-Based Attacks
Recognition in IoT-Oriented Industrial Control Systems**

Author: Akash Deep Bhardwaj, Fadi Al-Turman

Year: 2020

Description: Industrial Control Systems monitor, automate, and operate complex infrastructure and processes that integrate into critical industrial sectors that affect our daily lives. With the advent of networking and automation, these systems have moved from being dedicated and independent to centralized corporate infrastructure. While this has facilitated the monitoring and overall management using traditional detection methods, Web Application Firewalls or Intrusion Detection Systems has exposed the networks subjecting them to Behavior-based cybersecurity attacks. Such attacks alter the control flow and processes and have the malicious ability to alter the functioning of these systems altogether. This research focuses on the use of process analytics to detect attacks in the industrial control infrastructure systems and compares the effectiveness of signature-based detection methods. The proposed work presents a pattern recognition algorithm aptly named as “Capturing-the-Invisible (CTI)” to find the hidden process in industrial control device logs and detect Behavior-based attacks being performed in real-time.

2)Title: Image Forensics Based on Lighting Estimation

Author: Manoj Kumar, Anuj Rani and Sangeet Srivastava

Year: 2019

Description: Computer generated images are assumed to be a key part in each person's life in this era of information technology, where individuals effectively inhabit the advertisements, magazines, websites, televisions and many more. At the point when digital images played their role, the event of violations in terms of misrepresentation of information, use of their wrong doings winds up and also becomes easier with the help of image editing application programs. To be legitimate, if anyone does wrong anything then the proposed method can be used for a correct identification of the forgery and the imitations in the digital images. In existing techniques, researchers have suggested most well-known types of digital photographic manipulations based on source, meta-data, image copying, splicing and many more. The proposed approach is inspired by physics-based techniques and requires less human involvement. The presented approach works for images having any type of objects present in the scene, i.e. not only limited to human faces and selection of same intensity regions of the image. By assessing the lighting parameters, the proposed technique identifies the manipulated object and returns angle of incidence w.r.t light source direction. The demonstrated result produces forgery recognition rate of 92% on an image dataset comprising of various types of manipulated images.

3)Title: Secure attribute-based data sharing for resource-limited users in cloud computing

Author: JinLi^a, YinghuiZhang^{bcd}

Year: 2018

Description: Data sharing becomes an exceptionally attractive service supplied by

cloud computing platforms because of its convenience and economy. As a potential technique for realizing fine-grained data sharing, attribute-based encryption (ABE) has drawn wide attentions. However, most of the existing ABE solutions suffer from the disadvantages of high computation overhead and weak data security, which has severely impeded resource-constrained mobile devices to customize the service. The problem of simultaneously achieving fine-grained access, high-efficiency on the data owner's side, and standard data confidentiality of cloud data sharing actually still remains unresolved. This paper addresses this challenging issue by proposing a new attribute-based data sharing scheme suitable for resource-limited mobile users in cloud computing. The proposed scheme eliminates a majority of the computation task by adding system public parameters besides moving partial encryption computation offline. In addition, a public ciphertext test phase is performed before the decryption phase, which eliminates most of computation overhead due to illegitimate ciphertexts. For the sake of data security, a Chameleon hash function is used to generate an immediate ciphertext, which will be blinded by the offline ciphertexts to obtain the final online ciphertexts. The proposed scheme is proven secure against adaptively chosen-ciphertext attacks, which is widely recognized as a standard security notion. Extensive performance analysis indicates that the proposed scheme is secure and efficient.

4)Title: Ensuring attribute privacy protection and fast decryption for outsourced data security in mobile cloud computing

Author: Yinghui Zhang, Xiaofeng Chen

Year: 2017

Description: Although many users outsource their various data to clouds, data security and privacy concerns are still the biggest obstacles that hamper the widespread adoption of cloud computing. Anonymous attribute-based encryption (anonymous ABE) enables fine-grained access control over cloud storage and preserves receivers' attribute privacy by hiding attribute information in ciphertexts. However, in existing anonymous ABE work, a user knows whether attributes and a

hidden policy match or not only after repeating decryption attempts. And, each decryption usually requires many pairings and the computation overhead grows with the complexity of the access formula. Hence, existing schemes suffer a severe efficiency drawback and are not suitable for mobile cloud computing where users may be resource-constrained. In this paper, we propose a novel technique called match-then-decrypt, in which a matching phase is additionally introduced before the decryption phase. This technique works by computing special components in ciphertexts, which are used to perform the test that if the attribute private key matches the hidden access policy in ciphertexts without decryption. For the sake of fast decryption, special attribute secret key components are generated which allow aggregation of pairings during decryption. We propose a basic anonymous ABE construction, and then obtain a security-enhanced extension based on strongly existentially unforgeable one-time signatures. In the proposed constructions, the computation cost of an attribute matching test is less than one decryption operation, which only needs small and constant number of pairings. Formal security analysis and performance comparisons indicate that the proposed solutions simultaneously ensure attribute privacy and improve decryption efficiency for outsourced data storage in mobile cloud computing.

5)Title: DivORAM: Towards a practical oblivious RAM with variable block size

Author: ZheliLiu^a, YanyuHuang^a

Year: 2018

Description: Oblivious RAM (ORAM) is important for applications that require hiding access patterns. Many ORAM schemes have been proposed but most of them support only storing blocks of the same size. For the variable length data blocks, they usually fill them up to the same length before uploading, which leads to an increase in storage space and network bandwidth usage. To develop the first practical ORAM with variable block size, we proposed the "DivORAM" by remodeling the tree-based ORAM structure. It employs an additively homomorphic encryption scheme

(Damgård–Jurik cryptosystem) executing at the server side to save the client computing overhead and the network bandwidth cost. As a result, it saves network bandwidth 30% comparing with Ring ORAM and 40% comparing with HIRB ORAM. Experiment results show that the response time of DivORAM is 10 × improved over Ring ORAM for practical parameters.

2.2 OPEN PROBLEMS IN EXISTING SYSTEM

ILLUMINATION

The slight change in lighting conditions cause a significant challenge for automated face recognition and can have a significant impact on its results. If the illumination tends to vary, the same individual gets captured with the same sensor and with an almost identical facial expression and pose, the results that emerge may appear quite different.

POSE

Facial Recognition systems are highly sensitive to pose variations. The pose of a face varies when the head movement and viewing angle of the person changes. The movements of head or differing POV of a camera can invariably cause changes in face appearance and generate intra-class variations making automated face recognition rates drop drastically. It becomes a challenge to identify the real face when the rotation angle goes higher. It may result in faulty recognition or no recognition if the database only has the frontal view of the face

RESOLUTION

The minimum resolution for any standard image should be 16*16. The picture with the resolution less than 16*16 is called the low-resolution image. These low-resolution images can be found through small scale standalone cameras like CCTV cameras in streets, ATM cameras, supermarket security cameras. These cameras can capture a small part of the human face area and as the camera is not very close to face, they can only capture the face region of less than 16*16. Such a low-

resolution image doesn't provide much information as most of them are lost. It can be a big challenge in the process of recognizing the faces.

AGEING

Face appearance/texture changes over a period and reflect as ageing, which is yet another challenge in facial recognition system. With the increasing age, the human face features, shapes/lines, and other aspects also change. It is done for visual observation and image retrieval after a long period.

For accuracy checking, the dataset for a different age group of people over a period is calculated. Here, the recognition process depends on feature extraction, basic features like wrinkles, marks, eyebrows, hairstyles, etc.

PRIVACY

Privacy issues exist with facial recognition technology because of the ability to identify and track whereabouts, which may constitute an invasion of rights. Facial recognition technology can track individual which makes people uneasy. In addition, significant data breaches are all too common these days and the personal information that facial recognition software collects are not immune

However, there are two major cons to this technology: bias and inaccuracy that unknowingly creep in. The bias is based on race, gender, age group, education, or some other demographic factor. With larger data sets, the issue gets magnified even further and might have meaningless or hollow implications. Another reason for doubt that bothers many is that facial configurations cannot be matched to fingerprints and hence are not sacrosanct when conveying an emotion. It varies from person to person, and also, humans are quite capable of hiding their emotions or suppressing a dominant emotion with another emotion

CHAPTER 3

EMOTION DETECTION PROCESS AND APPLICATIONS

3.1 MODULES IN FERDL:

FACE CAPTURING MODULE

In this phase we are capturing the faces of people for future processing using web camera or external capturing the image none of the process can be done and there is no chance of detecting the emotions

PREPROCESSING MODULE

After capturing images, we will preprocess the images. In this phase we will convert the color images to gray scale images.

TRAINING MODULE

In this phase we will prepare a dataset that is binary array of all captured images. The captured images will store in .YML file, which stores all facial data

FACE RECOGNITION MODULE

The initial step of face recognition process is to train the facial data to the host system. Using the web camera of the computer system, 60 distinct images of the face is taken. In this module we will recognize the faces using LBPH algorithm. LBPH is an acronym for local binary pattern histogram. It will recognize the faces with the face ID and NAME that are all previously stored.

EXPRESSION RECOGNITION MODULE

Facial expression recognition software is a technology which uses biometric markers to detect emotions in human faces. It extracts and analyzes information from an image it is able to deliver unfiltered, unbiased, emotional response or data

3.2 FRAMEWORK OVERVIEW:

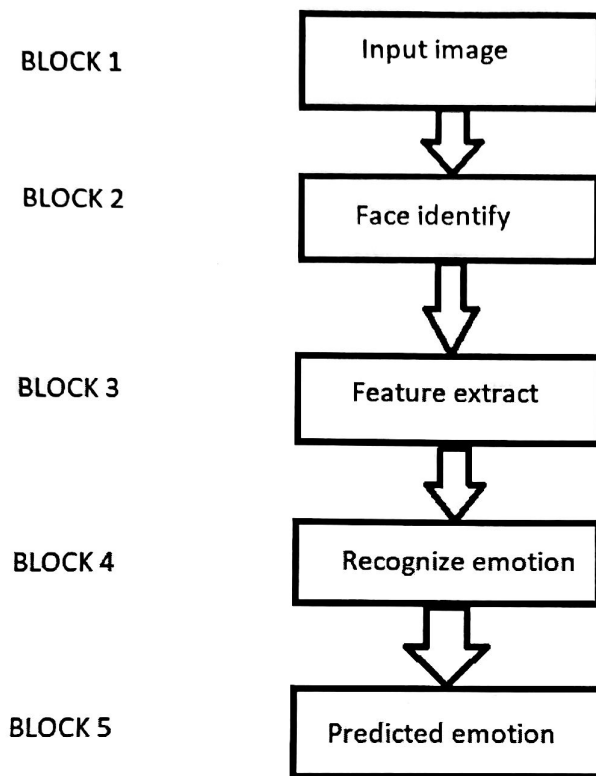


Fig:3.2.1 IMAGE RECOGNITION FRAMEWORK

CNN is that the hottest method of examining figures. The proposed methodology adopted within the present project work is depicted. The proposed method consists of two levels, the primary level recommended is to spot face, want to extract emotions from a picture. The next section finds the face characteristic such as eyes, mouth etc.

Hence the scenario might have serious repercussions where serious decision making is involved or if it affects people's everyday lives, rights, or opportunities. This can be deduced easily based on the increasing application of the technology that this technology is here to stay and prove its worth.

To make an efficient system that will be able to determine emotion from frontal face image is the main goal. In detail explanation of each block is given below:

- Block 1 gives the information about input image; this is collection of databases
- Block 2 gives the face identify; this will identify the face in the image.
- Block 3 gives the feature extract, facial features (like eyes and mouth) can be extracted.
- Block 4 gives the emotion recognition; recognition of the emotion can be done for the input image.
- Block 5 gives the emotion which has been detected from all the levels

3.3 SOFTWARE REQUIREMENTS IN FERDL

- a) IDLE (Python 3.6 64-bit)
- b) Java Development Kit
- c) OpenCV, Keras, TensorFlow and Tkinter libraries.

FACE DETECTION:

Face detection is a significant phase of FER. A proficient automated system can be developed for recognizing the face region in static or video image spontaneously. A face region is detected in the image sequence using facial features such as edge, skin color, texture and face muscle motion

FEATURE EXTRACTION:

After face detection, the next step in FER is feature extraction. The main aim of facial feature extraction is to extract an effective and efficient representation of facial components without any loss of face information. Geometric-based and

appearance-based features are the two feature extraction techniques classified based on facial motion and deformation of facial features. The input image may be either a static image or image sequence. Based on the input image, a suitable facial feature extraction algorithm is applied to extract either the local or global or hybrid features. The extracted features are considerably reduced in size, which is given as input to the classifier and that significantly helps the classifier to make the decision easier in identifying and recognizing the facial expression.

EXPRESSION CLASSIFICATION:

Classification is a supervised learning model to predict the output based on the observed values. In pattern recognition, the classifier plays a vital part either to draw a conclusion or to predict the class labels of unseen data based on trained images. Many classification algorithms were developed to achieve the highest recognition rate. Indeed, all the approaches are not suitable for analyzing different facial feature extraction techniques.

3.4 KEY FRAME EXTRACTION FROM INPUT VIDEO:

FERC works with an image as well as video input. In case, when the input to the FERC is video, then the difference between respective frames is computed. The maximally stable frames occur whenever the intra-frame difference is zero. Then for all of these stable frames, a Canny edge detector was applied, and then the aggregated sum of white pixels was calculated. After comparing the aggregated sums for all stable frames, the frame with the maximum aggregated sum is selected because this frame has maximum details as per edges (more edges more details). This frame is then selected as an input to FERC. The logic behind choosing this image is that blurry images have minimum edges or no edges

3.5 USE CASES OF EMOTION DETECTION

Emotion detection technology has reached new milestones with the increase in computing power, highly sophisticated deep learning models, and the availability of data sets on social media platforms. Large amounts of visual data in video, audio,

and image formats have resulted in an enormous repository of materials to power emotion detection capabilities.

Here are some examples of emotion detection and recognition applications:

Self-driving vehicles—futuristic car designs incorporate sensors, such as microphones and cameras, to enable monitoring and real-time analysis of conditions within and outside the vehicle. Smart cars can use facial emotion recognition to warn drivers when they feel drowsy. Emotion detection technology can capture subtle facial expressions that other facial recognition technologies cannot.

Medical diagnostics—doctors can use speech analysis and micro-expression detections to identify symptoms of dementia and depression.

Personalized education—educational software prototypes can respond to a child's emotions to improve the learning experience. If the child is frustrated that the task is too difficult or too easy, the program can adjust the task to make it easier or more difficult.

Smart retail—retailers can use computer vision technology and AI-powered emotion detection in their stores to gauge a shopper's reaction and mood and collect demographic information.

Virtual assistants—future versions of popular virtual assistants will be able to leverage emotion detection capabilities to improve the service. For example, the virtual assistant can modify its language and tone according to the user's mood. For example, Apple has submitted a patent and plans to integrate this capability into Siri

3.6 METHODS AND APPROACHES

After years of scientific research, various automatic emotion recognition methods have emerged. Researchers propose and evaluate different approaches, utilizing technologies from multiple domains such as machine learning, signal processing, speech processing, and computer vision.

Naive Bayes were used for emotion recognition. Today, state of the art algorithms is based on deep learning. A few examples:

Deep convolutional neural network (CNN) that extracts facial landmarks from data, reduces images to 48×48 pixels, and runs them through two convolution-pooling layers (Mollahosseini et al., 2016)

CNN for detecting facial movements using the concept of Action Units (AU), with two convolutional layers followed by a max pooling layer, and two fully connected layers that predict the number of AUs found in the face image (Mohammadpour et al., 2017)

Using three CNNs with the same architecture, each detecting a separate part of the face – eyebrow, eye, and mouth. The output is combined into an “iconic face” which is fed into another CNN to detect emotions (Yolcu et al., 2019)

3.7 6 TYPES OF DATASETS

Belfast Database

This database includes recordings of responses to lab-based emotional induction tasks—it contains data about self-reported emotions, intensity, sex, and task passivity or activeness. The Belfast database lets researchers compare the expressions of people from different cultures

SEMAINE

SEMAINE is a video dataset containing audiovisual data capturing the interactions between humans and one of four personalities (happy, gloomy, angry, and pragmatic). The video clips have annotations that rate epistemic state pairs like thoughtful, interested, certain, agreement, and concentration.

The Interactive Emotional Dyadic Motion Capture Dataset (IEMOCAP)

IEMOCAP contains 151 dialogue videos with two speakers each (302 in total). Annotations identify nine emotions—anger, fear, excitement, sadness, surprise, frustration, happiness, disappointment, and neutral. It also detects dominance and

valence.

DEAP

- a) This dataset contains EEG, ECG, and face videos with emotion annotations. It has two parts:
- b) Ratings from online self-assessments—volunteers rated 120 music video extracts by arousal, dominance, and valence.
- c) Face videos and physiological recordings—volunteers rated 40 music videos while undergoing EEGs or having their faces recorded.

Multimodal Emotion Lines Dataset (MELD)

MELD extends the Emotion Lines dataset containing dialogue instances, adding visual and auditory modalities. It has over 13,000 utterances from 1,400 dialogues from the TV show *Friends*. Each utterance has a label classifying the emotion—anger, disgust, fear, neutral, joy, sadness, or surprise—alongside negative, positive, or neutral annotations.

Affect Net

Affect Net is a database of facial expressions that was created to facilitate research on automatic facial expression recognition. It contains more than 1 million annotated images of facial expressions, with labels indicating the presence and intensity of various emotional states, such as happiness, sadness, anger, and fear.

3.8 MACHINE LEARNING'S APPLICATION

- Compared to conventional ER, ER by machine learning can greatly enhance a dynamic accuracy.

- Contemporary machine learning methods encompass logistic regression (LR), decision tree, artificial
- neural network (ANN), naive Bayes, random forest, etc. [3]. There are several machine learning
- models in the field of emotion recognition.
- Convolutional neural network (CNN)
- CNN has been proven to be a valid and reliable machine learning approach in recognizing images,
- and therefore, has been frequently used in recognizing EEG images [3].
- Highlights in Science, Engineering and Technology MISBP 2022
- Volume 36 (2023)
- 405
- Li et al deploy and train hierarchical CNN(HCNN) with features to recognize positive, negative,
- and neutral moods. Results have proved CNN a potent candidate on beta and gamma waves [11].
- Deep belief network (DBN)
- DBN consists of stacked Restricted Boltzmann Machine (RBMs) and can be used in supervised
- and unsupervised learning tasks [12].
- The emo F-BVP multimodal database is studied with 23 different emotions generated by actors.
- Four DBN models are used for analysis and contribute to the improved recognition results than
- current state of art methods. A convolutional deep belief network (CDBN) is also proposed,
- presenting higher accuracy (79.2%) in subtle emotion recognition.
- Recurrent neural network (RNN)
- RNN is another advanced deep learning algorithm which perform better in time

series data. Long

- short-term memory (LSTM) belongs to RNN and has its benefits in memory processing.
- sequential contexts.
- Salma et al apply this method to DEAP data set for classifying low/high valence, arousal, and
- liking status from EEG features. Results have shown an average accuracy of 85.65% in arousal, 85.45%
- in valence, and 87.99% in liking, surpassing previously proposed methods.
- Artificial neural network (ANN)
- Besides methods described above, artificial neural network is also regarded as a promising
- direction for accurate emotion recognition. ANN is composed of connected artificial neurons to
- imitate natural brain. It can play a part in handling complex situations. Studies using ANN have
- highlighted its reliable performance, some reaching over 90% of accuracy in emotion recognition and
- outperforming other methods.

CHAPTER 4

DESCRIPTION OF PROPOSED SYSTEM

4.1 SELECTED METHODOLOGY OR PROCESS MODEL

Convolutional Neural Networks are a form of deep neural network that is used for computer vision and visual image processing. However, conventional algorithms face certain serious issues or questions, such as luminous variance and location variance, etc. The approach to addressing the problems of conventional methods is to apply the CNN algorithm to the classification of emotions. In contrast, our model is sequentially structured. We recognize that Sequence Modeling could model, analyze, make predictions or produce some form of sequential data. In comparison to the traditional form, the algorithm's major distinctions are:

- Feature extraction with LBP
- With the use of Local Binary Pattern (LBP), image feature map can be extracted, as the feature extractors are generated during the training and testing process.
- Differences of mathematical mode
- Typically, the linear classifier is classified by linear transformation. This is commonly referred to as the traditional form. In contrast, to discern variations in the classification process, CNN and other deep learning algorithms usually incorporate linear conversion with nonlinear features such as sigmoid and rectified linear unit functions (ReLU). We also take the advantages of strides and padding parameters in our model architecture.
- The traditional approach usually conducts only one layer of an operation via the linear classifier: SVM has just one weight set, for instance. However, in the course of classification, CNN and other deep learning algorithms perform several layers of operation. As a two-dimensional array, CNN adopts input

4.1.1 PROCESS MODEL FOR EMOTION DETECTION

The Component Process Model is a well-established framework describing an emotion as a dynamic process with five highly interrelated components: cognitive appraisal, expression, motivation, physiology and feeling. Yet, few empirical studies have systematically investigated discrete emotions through this full multi-componential view. We therefore elicited various emotions during movie watching and measured their manifestations across these components. Our goal was to investigate the relationship between physiological measures and the theoretically defined components, as well as to determine whether discrete emotions could be predicted from the multicomponent response patterns. By deploying a data-driven computational approach based on multivariate pattern classification, our results suggest that physiological features are encoded within each component, supporting the hypothesis of a synchronized recruitment during an emotion episode. Overall, while emotion prediction was higher when classifiers were trained with all five components, a model without physiology features did not significantly reduce the performance. The findings therefore support a description of emotion as a multicomponent process, in which emotion recognition requires the integration of all the components. However, they also indicate that physiology per se is the least significant predictor for emotion classification among these five components.

In sum, there is still no unanimous conclusion about distinguishable patterns of activation in ANS, due to the difficulty to identify and associate reliable response patterns to discrete emotions. Consequently, the debate is not closed concerning the functionality of physiology during an emotional experience. Based on the CPM model, physiology is involved in shaping emotion and can contribute to differentiating emotion. However, while relevant, we hypothesize that the use of physiology alone is limited in discriminating emotion but could be better understood if integrated with the other major components of emotion. Therefore, to provide further insights about the contribution of physiology in emotion differentiation, we propose here to examine how a full componential model can account for the multiple and concomitant changes in physiological and behavioral measures observed during emotion elicitation. In addition, we examine the added information by each component and hypothesize that

considering the synchronized changes in all components, the information in each component is already encoded in the other components. To the best of our knowledge, the present work represents one of the first attempts to investigate the componential theory by explicitly considering a combination of multiple, theoretically defined, emotional processes that occur in response to naturalistic emotional events (from cinematic film excerpts). By deploying a data-driven computational approach based on multivariate pattern classification, we aim at performing detailed analyses of physiological data in order to distinguish and predict the engagement of different emotion components across a wide range of eliciting events. On the grounds of such multicomponent response patterns, we also aim at determining to what extent discrete emotion categories can be predicted from information provided by these components, and what is the contribution of each component in such predictions. We hypothesize that a multicomponent account, as proposed by the CPM (Scherer, 1984, 2009), may allow us to capture the variability of physiological activity during emotional episodes, as well as their differentiation across major categories of emotions

The whole experiment consisted of four sessions scheduled on different days. Each session was divided into two parts, fMRI experiment and behavioral experiment, lasting for about 1 and 2 h, respectively. In the current study we focus only on the behavioral analysis and will not use the fMRI data. Stimuli presentation and assessment were controlled using Psychtoolbox-3, an interface between MATLAB and computer hardware.

During the fMRI experiment, participants were engaged in an emotion elicitation procedure using our 40 emotional film excerpts. No explicit task was required during this phase. They were simply instructed to let themselves feel and express emotions freely rather than controlling feelings and thoughts because of the experiment environment. Movies were presented inside the MRI scanner on an LCD screen through a mirror mounted on the head coil. The audio stream was transmitted through MRI-compatible earphones. Each session was composed of 10 separate runs, each presenting a film clip preceded by a 5-seconds instruction screen warning about the imminent next display and followed by a 30-seconds washout periods introduced as a low-level perceptual control baseline for the fMRI analysis (not analyzed here). Moreover, a session consisted of a pseudo-random choice of 10 unique film clips with

considering the synchronized changes in all components, the information in each component is already encoded in the other components. To the best of our knowledge, the present work represents one of the first attempts to investigate the componential theory by explicitly considering a combination of multiple, theoretically defined, emotional processes that occur in response to naturalistic emotional events (from cinematic film excerpts). By deploying a data-driven computational approach based on multivariate pattern classification, we aim at performing detailed analyses of physiological data in order to distinguish and predict the engagement of different emotion components across a wide range of eliciting events. On the grounds of such multicomponent response patterns, we also aim at determining to what extent discrete emotion categories can be predicted from information provided by these components, and what is the contribution of each component in such predictions. We hypothesize that a multicomponent account, as proposed by the CPM (Scherer, 1984, 2009), may allow us to capture the variability of physiological activity during emotional episodes, as well as their differentiation across major categories of emotions

The whole experiment consisted of four sessions scheduled on different days. Each session was divided into two parts, fMRI experiment and behavioral experiment, lasting for about 1 and 2 h, respectively. In the current study we focus only on the behavioral analysis and will not use the fMRI data. Stimuli presentation and assessment were controlled using Psychtoolbox-3, an interface between MATLAB and computer hardware.

During the fMRI experiment, participants were engaged in an emotion elicitation procedure using our 40 emotional film excerpts. No explicit task was required during this phase. They were simply instructed to let themselves feel and express emotions freely rather than controlling feelings and thoughts because of the experiment environment. Movies were presented inside the MRI scanner on an LCD screen through a mirror mounted on the head coil. The audio stream was transmitted through MRI-compatible earphones. Each session was composed of 10 separate runs, each presenting a film clip preceded by a 5-seconds instruction screen warning about the imminent next display and followed by a 30-seconds washout periods introduced as a low-level perceptual control baseline for the fMRI analysis (not analyzed here). Moreover, a session consisted of a pseudo-random choice of 10 unique film clips with

high ratings on at least one of the 10 different pre-labeled discrete emotion categories (fear, anxiety, anger, joy, sadness, satisfaction, surprise, love, disgust, calm). This permitted to engage potentially different component processes in every session. To avoid any order effect, the presentation of all stimuli was counterbalanced.

The behavioral experiment was performed at the end of each fMRI session, in a separate room. Participants were let alone with no imposed time constraints to complete the assessment. They were asked to rate their feelings, thoughts, or emotions evoked during the first viewing of the film clips and advised not to report what might be expected to feel in general when watching such kinds of events. To achieve the emotion evaluation, the 10 film excerpts seen in the preceding session were presented on a laptop computer with LCD screen and headphones. However, the previously selected highly emotional segments (see “stimuli selection” above) were now explicitly highlighted in each film excerpts by a red frame surrounding the visual display. In order to ensure that emotion assessment corresponded to a single event and not the entire clip, the ratings were required right after each segment by pausing the clips. The assessment involved a subset of Core GRID instrument (Fontaine et al., 2013), which is to date the most comprehensive attempt for multi-componential measurement in emotion. The set of 32 items (see Table 1) had been pre-selected based on their applicability to the emotion elicitation scenario with movies, rather than according to an active first-person involvement in an event. Among our set of Core GRID items, 9 were related to the appraisal component, 6 to the expression component, 7 to the motivation component, 6 to the feeling component, and 4 to the bodily component. Participants had to indicate how much they considered that the description of the Core GRID items correctly represented what they felt in response to the highlighted segment, using a 7-level Likert scale with 1 for “not at all” and 7 for “strongly.”

PHYSIOLOGICAL DATA ACQUISITION:

Several physiological measures were collected during the first part of each session in the MRI scanner, including heart rate, respiration rate, and electrodermal activity. All the measures were acquired continuously throughout the whole scanning time. The data were first recorded with a 5,000 Hz sampling rate using the MP150

Biopic Systems software (Santa Barbara, CA), before being pre-processed with Acknowledge 4.2 and MATLAB 2012b.

Heart rate (HR) was recorded with a photoplethysmogram amplifier module (PPG100C). This single channel amplifier designed for indirect measurement of blood pressure was coupled to a TSD200-MRI photoplethysmogram transducer fixed on the index finger of the left hand. Recording artifacts and signal losses were corrected using endpoint function from Acknowledge, which interpolates the values of a selected impaired measure portion. Secondly, the pulse signal was exported to MATLAB and down sampled to 120 Hz. To remove scanner artifacts, a comb-pass filter was applied at 17.5 Hz. The pulse signal was then filtered with a band-pass filter between 1 and 40 Hz. Subsequently, the instantaneous heart rate was computed by identifying the peaks in the pulse signal, calculating the time intervals between them and converting this distance into beats per minute (BPM). The standard heart rate in humans goes from 60 to 100 bpm at rest. Hence, it was considered that a rate above 100 bpm was unlikely and the minimum distance between peaks will not exceed this limit. This automatic identification was manually verified by adding, changing, or removing the detected peaks and possible outliers.

Respiration rate (RR) was measured using a RSP100C respiration pneumogram amplifier module, designed specifically for recording respiration effort. This differential amplifier worked with a TSD201 respiration transducer, which was attached with a belt around the upper chest near the level of maximum amplitude in order to measure thoracic expansion and contraction. Using a similar procedure as for HR preprocessing, the connect endpoint function of acknowledge was first employed to correct manually the artifacts and losses of signal. After exporting the raw signal to MATLAB, it was down sampled to 120 Hz and then filtered with a band pass filter fixed between 0.05 and 1 Hz. Lastly, the signal was converted to breaths per minute using the same procedure as above. The standard respiration rate in human goes from 12 to 20 breaths per minute at rest. Since participants were performing a task inside a scanner which could be an unusual environment, the higher maximum rate was increased at 35 cycles per minute. Therefore, it was estimated that a rate above 35 was unlikely and the minimum distance between peaks will not exceed this limit. Again, this information was used in the automatic detection of the signal peaks. The respiration rate was then manually verified by looking at the detected signal peaks

and corrected, with outliers being removed when it was necessary.

Electrodermal activity (EDA) was registered using an EDA100C electrodermal activity amplifier module, a single-channel, high-gain, differential amplifier designed to measure skin conductance via the constant voltage technique. The EDA100C was connected to Adult ECG Clear trace 2 LT electrodes. Electrodes were placed on the index and the median fingers of the participants left hand. Following the manual correction of artifacts and losses of signal with the connect endpoint function on acknowledge, the raw signal was exported to MATLAB. Similar to the two other physiological signals, the EDA signal was down sampled to 120 Hz. This signal, recorded by BIOPAC in microSiemes (μS), was then filtered with a 1 Hz low pass filter. An IIR (infinite impulse response) high-pass filter fixed at 0.05 Hz was applied to derive the Skin Conductance Responses (phasic component of EDA) representing the rapidly changing peaks, while a FIR (finite impulse response) low-pass filter fixed at 0.05 Hz was applied to derive the Skin Conductance Levels (tonic component of EDA) corresponding to the smooth underlying slowly-changing levels (acknowledge 4 Software Guide, 2011).

FEATURES AND NORMALIZATION

MATLAB was used to select physiological values during the 12-s duration of high emotional segments. From these values, the means, variances, and ranges of each physiological signals (HR, RR, phasic and tonic EDA) were calculated. We chose to focus specifically on the mean and variance of these physiological signals, as these are the most reliable and frequently reported features in studies associating discrete emotions and physiological responses (Kreibig, 2010). For HR and RR measures, respectively 4 and 17 responses during highly emotional segments had to be removed in one participant due to a corrupted signal, potentially induced by movements. For EDA, 267 values had to be removed due to temporary losses of signal, resulting in flat and useless measures. In particular, EDA responses of two subjects were completely removed as the EDA sensor could not capture their response. In order to handle the missing values in the physiological data, dropouts were replaced by mean value of the whole session during which the signal loss has happened (i.e., missing value imputation). Furthermore, the variance in physiological

responses could be very large and different across participants. Because it was particularly important to reduce such variability in order to avoid inter-individual biases, all physiological measures were normalized within-subject using RStudio (1.1.383). To achieve this, standardized z-scores were calculated from the physiological data during the 4 sessions of each participant. Regarding responses collected for the 32 CoreGRID items for each high emotional segment, a within-subject normalization into z-score was also performed. These normalized behavioral data and the discrete emotion labels selected by the participants for each emotional segment were incorporated to the related physiological measures. In the end, for each of the 119 emotional segments, we obtained a set of observations including 32 standardized CoreGRID items and 1 discrete emotion label, as well as 8 standardized physiological values calculated offline. However, the final dataset included 19 participants who attended all 4 sessions ($19 \times 119 = 2,261$), while 1 participant completed 2 sessions out of 4 ($1 \times 55 = 55$). Also, for 11 participants, the assessment of one of the emotional segments did not get recorded due to a technical issue. Therefore, in total, there were ($2,261 + 55 - 11 =$) 2,305 sets of observations instead of the possible maximum of ($20 \times 119 =$) 2,380 (~3% of points loss)

PREDICTIVE ANALYSIS

To investigate the relationship between physiology and the component model descriptors, two analyses were performed. First, we examined whether the physiology measures allowed predicting component model descriptors and vice versa. Second, we assessed whether distinct features from the componential model allowed predicting discrete emotion categories and compared the value of different components for this prediction. For both analyses, multivariate pattern classifications using machine learning algorithms were undertaken to predict the variables of interest. Linear and non-linear classifiers including Logistic Regression (LR) and Support Vector Machine (SVM) with different kernels (linear, radial basis function, polynomial and sigmoid) were applied. All analyses were carried out using the RStudio statistical software, Version 1.1.383. Logistic regressions were conducted with the “caret” package, Version 6.0 and multinomial logistic regressions with the “nnet” package, Version 7.3. The binary and multiclass classifications using Support Vector Machine were conducted with the “e1071” package, Version 1.7.

First, the CoreGRID items were used as predictor variables to predict the dependent variable, which was either the mean or the variance of each physiological measure. To enable such analyses and to simplify the computational problem, the scores of the dependent variable were converted into two classes of “High” and “Low” using the median value across all the participants as a cutoff threshold. LR and SVM with linear and non-linear kernels using 10-fold cross-validation were applied. To guarantee test and training independence, each participant's assessment was included in either a test set or a training set. Conversely, similar analyses were carried out to determine whether physiology measures could encode the component model descriptors, but now using the physiology measures as independent variables in an attempt to predict the ratings of each CoreGRID item as either above or below the median.

Secondly, to examine the relationship between the component process model and discrete emotion types, multiclass classifications using SVM were performed on different combinations of CoreGRID items and physiological measures in order to predict specific emotion categories as labeled by the participants. Given the large number of classes and limited number of samples per class with too many predictors, a leave-one-subject-out cross-validation was used to guarantee a complete independence between the training and testing datasets.

For all analyses, we used a grid search method to optimize the parameters, but no significant improvement was observed, so the default parameters were kept. Moreover, SVM with radial basis function (RBF) kernel outperformed LR and other SVM models. Therefore, we will only report the result from SVM with RBF kernel.

It should be mentioned that the classes used for binary classifications were pretty balanced since they were defined based on the median value of the dependent variable, resulting in a distribution close to 50–50 split. However, in the case of multi-class classification, although the number of movie clips for each pre-labeled emotion category was balanced, the final dataset was not since we used the subjectively experienced emotions reported by each participant as emotional labels. This imbalance may have slightly affected the classification performance for some under- or over-represented categories. To account for an effect of class distribution, we reported the chance level in all comparisons as well as the confusion matrix.

Assuming that physiological responses are intertwined with all components of emotion, we expected that ratings on the 32 CoreGRID features would carry information sufficient to predict corresponding physiological changes. Effectively, SVM classifications provided prediction accuracies significantly better than the chance level. However, information from the CoreGRID items did not allow a high accuracy, even though prediction was simplified by being restricted to a binary distribution. This modest accuracy may be explained by a great variability across participants, which could reduce the generalizability of classifiers when they were applied to all individuals rather than within subject. It might also reflect heterogeneity in intra-individual physiological responses among emotions with similar componential patterns. In parallel, the opposite approach to predict ratings of CoreGRID features based on the means and variances of physiological responses also yielded a performance significantly higher than chance level but still relatively low. Because each CoreGRID item focuses on quite specific behavioral features, it is however not surprising that the sole use of physiology would be insufficient to precisely determine the ratings.

4.1.2 FACIAL RECOGNITION WITH MUSIC RECOMMENDATION

Facial and emotion popularity may be completed the use of an expansion of features, which includes facial speech and even textual content. Among these functions, facial expressions are several the maximum popular, making it less difficult to attain a huge number of facial statistics (another tool for human recognition). Recently, with using deep studying and mainly Convolutional Neural Networks (CNs), many capabilities can be extracted and educated for a decent Viola Jones gadget. In an up-to-date form, music hints are brought into this method to lessen stress stages and exchange temper

4.2 ARCHITECTURE / OVERALL DESIGN OF PROPOSED SYSTEM

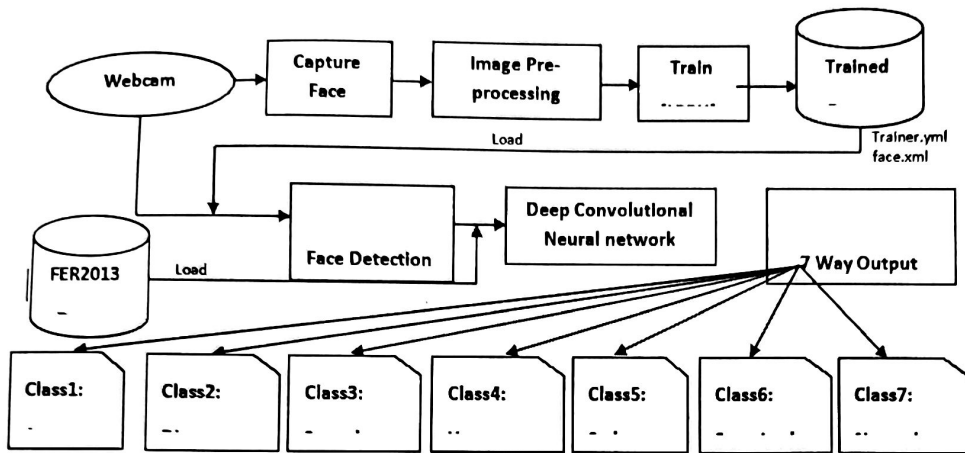


Fig 4.2.1: ARCHITECTURE OF FERDL

4.3 EMOTION RECOGNITION USING CNN

DATA PREPROCESSING

The column in pixel first is stored in a list format. Since computational complexity is high for computing pixel values in the range of (0-255), the data in pixel field is normalized to values between [0-1]. The face objects stored are reshaped and resized to the mentioned size of 48 X 48. The respective emotion labels and their respective pixel values are stored in objects. We use scikit-learns `train_test_split()` function to split the dataset into training and testing data. The test size being 0.2 meaning, 20% of data is for validation while 80% of the data will be trained.

IMAGE AUGMENTATION

More data is generated using the training set by applying transformations. It is required if the training set is not sufficient enough to learn representation. The image data is generated by transforming the actual training images by rotation, crop, shifts, shear, zoom, flip, reflection, normalization etc..

FEATURE EXTRACTION

Among all facial features, eye localization and detection is essential, from which locations of all other facial features are identified. However, the existing face recognition techniques failed to identify the exact person. In order to overcome this issue, Geometric Shaped Facial Feature Extraction for Face Recognition (GSF 2 EFR) is designed for identifying the exact person by finding the center and corners of the eye using eye detection and eye localization modules. An input facial image is given where the face map gets initialized and processed in the form of course-to-fine manner with two modules. In the first module, eye detector is used to detect the eye pattern using Gabor filter

DATA TRAINING

The 2- and 3-dimensional automated facial recognition techniques find more and more application possibilities. Aside from their original purpose – which is the identification or verification of persons in the safety and criminal-law domain – numerous fields of application have been added in the economic and research industries. With the help of facial recognition, robots can, for example, recognize human emotions and react accordingly. Online shops selling glasses can offer suitable frames for every user through facial recognition and the clients can virtually try them on. Through automated facial recognition, person-specific styling as well as make-up tips can be given online and new stylings can be visualized

VALIDATION

In the validation phase, various OpenCV functions and Keras functions have been used. Initially the video frame is stored in a video object. An LBP cascade classifier is used to detect facial region of interest. The image frame is converted into grayscale and resized and reshaped with the help of numpy. This resized image is fed to the

predictor which is loaded by `keras.models.load_model()` function. The max argument is output. A rectangle is drawn around the facial regions and the output is formatted above the rectangular box.

The convex hull have been computed, we determine the triplet formed by the angles on the sides of the two first centroids with the lowest y-coordinates and the absolute value of their difference. In the successful cases, the two centroids with the lowest y-coordinates are the two eyes. The set of triplets is further used as input to the k-means (with $k = 2$) clustering algorithm and the centroids of each class is extracted, one representing the class of the set of successful detections, and the other the set of incorrect detection

4.4 CONVOLUTION FILTER

The entire image is divided into overlapping 3×3 matrices, and then the corresponding 3×3 filter is convolved over each 3×3 matrix obtained from the image. The sliding and taking dot product operation is called 'convolution' and hence the name 'convolutional filter.' During the convolution, dot product of both 3×3 matrix is computed and stored at a corresponding location, e.g., (1,1) at the output. Once the entire output matrix is calculated, then this output is passed to the next layer of CNN for another round of convolution. The last layer of face feature extracting CNN is a simple perceptron, which tries to optimize values of scale factor and exponent depending upon deviation from the ground truth. The main advantage is good at detecting patterns and features in images, videos, and audio signals. Robust to translation, rotation, and scaling invariance. End-to-end training, no need for manual feature extraction. Can handle large amounts of data and achieve high accuracy. Computationally expensive to train and require a lot of memory. Can be prone to overfitting if not enough data or proper regularization is used. Requires large amounts of labeled data. Interpretability is limited, it is hard to understand what the network has learned. which is predominantly used to extract the feature from the grid-like matrix dataset. For example, visual datasets like images or videos where data patterns play an extensive role

CHAPTER 5

EXPERIMENTAL ANALYSIS AND EVALUATION

5.1 EXPERIMENTAL ENVIRONMENT

To test the pre-attentiveness and effectiveness of different features in Chernoff faces, we conducted a controlled experiment with twenty-four subjects. Display screens were created by a Java 1.1 application, which displayed Chernoff faces for a set amount of time. The experiments were run on a Sun Sparc 5.6 workstation. This experiment employed a two-factor, within objects design. The two factors, which were systematically varied to create the test conditions, were target feature and number of faces in the display. Four levels of number of faces were tested (5, 10, 25, or 50). The order of presentation of the different number of faces was randomized. Local Binary Pattern (LBP) is a term used to describe the texture properties of pictures on a local level. The major benefit of the Local binary pattern is its rotation and grey invariance. LBP is a basic technique for detecting characteristics in an image that is resistant to lighting fluctuations. It is a commonly utilized approach for feature extraction in various object identification and facial expression detection algorithms because of its simplicity and durability. The Local Binary Pattern is an effective texture classifier. The binary operator labels the pixels in the image by comparing the centre pixel value with the 3×3 neighbourhood of each pixel value to create an 8-bit binary number, which is then transformed to a decimal value. The binary-valued picture patch that results is used to create a local image descriptor³⁷. The following equation is used to calculate LBP code for a given pixel at (x_c, y_c) :

The accuracy of multi-class emotion detection systems has substantial room for improvement. The study in¹⁰ examined different types of classifiers on different feature types. The researchers extracted statistical features, power density, frequency-domain features, entropy, and wavelet energy. They chose 14 electrodes from the DEAP dataset and extracted up to 168 features for every subject. They applied KNN, SVM, and decision tree (DT) classifiers. The optimal result for valence and dominance was associated with the use of the KNN classifier with statistical EEG features. Specifically, valence amounted to 77.54%, and dominance amounted to 79%. For the same features, arousal was 78.88% using the SVM classifier.

5.2 STRUCTURE OF CNN

At first, we see a simple CNN template with several building blocks that we can easily understand and correlate to the proposed CNN model. Three types of layers make up a basic CNN, input, hidden, and output. The data enters the CNN via the input layer and then travels through many hidden levels before reaching the output layer. The network's prediction is reflected via the output layer. In terms of loss or error, the network's output is compared to the actual labels

5.3 RECOGNIZING FACIAL EMOTION FROM LIVE VIDEOS

- Recognizing Facial Emotions from Live Videos
- Like our processing of images for sentiment extraction, we will work on videos.
- Theoretically, a video is a combination of continuous image frames put together in motion.

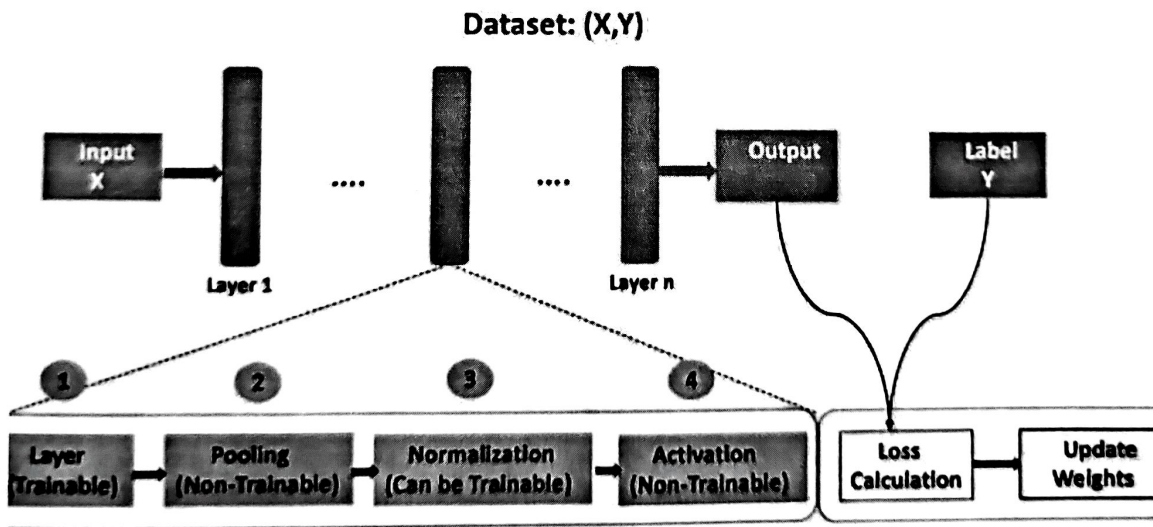


Fig 5.3.1: IMAGE FEATURE EXTRACTION MODEL

So essentially, the working of any algorithm for both videos and images is the same. The only additional step in processing videos is splitting the video into all of its

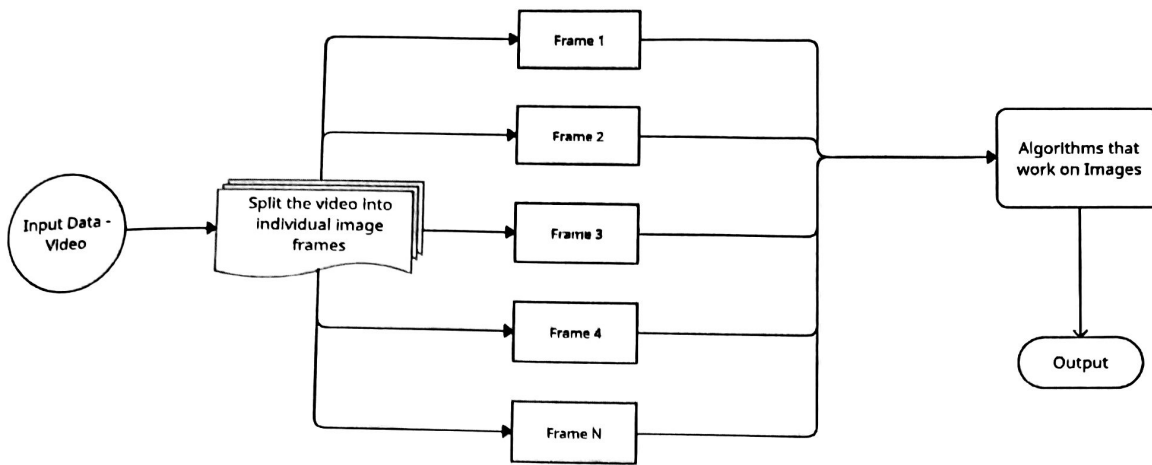


Fig 5.3.2: PROCESS MODEL OF FER

individual frames and then applying the image processing algorithms to it.

The flow of Logic: Although the underlying algorithm is similar for both images and videos, there are a few key changes that we will follow for videos.

- 1) Video analyze (): This function is responsible for extracting the individual image frames from a video and then analyze those independently.
- 2) Each frame analyzed by this function is stored as a separate image by the algorithm in the root directory folder where the code is running. Also, this function later creates a replica of the original video by placing a box around the face and showing live emotions within the video.
- 3) We then create a Pandas DataFrame from these analyzed values and plot this dataframe using matplotlib. In this plot, we can see every emotion plotted against time.
- 4) We can further analyze this dataframe by taking individual emotion values that were recognized by the model and finding which sentiment was dominant across the entire video.
- 5) This way, we can work on videos by extracting individual image frames and analyzing them. This process is displayed in the diagram below that shows how an additional step gets added for processing videos. We will be seeing this implementation in the section below.

5.4 IMPLEMENTATION

```
from keras. layers import Activation, Convolution2D, Dropout, Conv2D
from keras. layers import AveragePooling2D, Batch Normalization
from keras. layers import GlobalAveragePooling2D
from keras. models import Sequential
from keras. layers import Flatten
from keras. models import Model
from keras. layers import Input
from keras. layers import MaxPooling2D
from keras. layers import SeparableConv2D
from keras import layers
from keras. regularizers import l2
```

Unlike a function, though, layers maintain a state, updated when the layer receives data during training, and stored in `layer.weights`:

```
model.add(Convolution2D(filters=256, kernel_size=(3, 3), padding='same'))
    model.add (BatchNormalization())
    model.add      (Convolution2D(filters=num_classes,      kernel_size=(3,      3),
padding='same'))
    model.add(GlobalAveragePooling2D())
    model.add(Activation('softmax',name='predictions'))
    return model
```

This parameter is an integer or tuple/list of 2 integers, specifying the “step” of the convolution along with the height and width of the input volume.

Its default value is always set to (1, 1) which means that the given Conv2D filter is applied to the current location of the input volume and the given filter takes a 1-pixel step to the right and again the filter is applied to the input volume and it is performed until we reach the far-right border of the volume in which we are moving our filter

```
model.add(Convolution2D(filters=64, kernel_size=(3, 3), padding='same'))
    model.add(BatchNormalization())
    model.add(Convolution2D(filters=64, kernel_size=(3, 3),
        strides=(2, 2), padding='same'))
    model.add(BatchNormalization())
    model.add(Activation('relu'))
    model.add(Dropout(.25))

    model.add(Convolution2D(filters=64, kernel_size=(1, 1), padding='same'))
```

```

model.add(BatchNormalization())
model.add(Convolution2D(filters=128, kernel_size=(3, 3),
                        strides=(2, 2), padding='same'))
model.add(BatchNormalization())
model.add(Activation('relu'))
model.add(Dropout(.25))

model.add(Convolution2D(filters=256, kernel_size=(1, 1), padding='same'))
model.add(BatchNormalization())
model.add(Convolution2D(filters=128, kernel_size=(3, 3),
                        strides=(2, 2), padding='same'))

model.add(Convolution2D(filters=256, kernel_size=(1, 1), padding='same'))
model.add(BatchNormalization())
model.add(Convolution2D(filters=num_classes, kernel_size=(3, 3),
                        strides=(2, 2), padding='same'))

```

Padding

- The padding parameter of the Keras Conv2D class can take one of two values: 'valid' or 'same'.
- Setting the value to "valid" parameter means that the input volume is not zero-padded and the spatial dimensions are allowed to reduce via the natural application of convolution

Data format

- This parameter of the Conv2D class can be either set to "channels last" or "channels first" value.
- The TensorFlow backend to Keras uses channels last ordering whereas the Theano backend uses channels first ordering.
- Usually, we are not going to touch this value as Keras as most of the times we will be using TensorFlow backend to Keras.
- It defaults to the image_data_format value found in your Keras config file at ~/.keras/Keras.json.

Dilation rate

- The dilation rate parameter of the Conv2D class is a 2-tuple of integers, which

controls the dilation rate for dilated convolution.

- The Dilated Convolution is the basic convolution applied to the input volume with defined gaps.
- You may use this parameter when working with higher resolution images and fine-grained details are important to you or when you are constructing a network with fewer parameters.
- activation

The activation parameter to the Conv2D class is simply a convenience parameter which allows you to supply a string, which specifies the name of the activation function you want to apply after performing the convolution

UNDERSTANDING THE CODE:

When adding the Conv2D layers using `Sequential.model.add ()` method, there are numerous parameters we can use which we have read about earlier in our blog.

The first parameter tells us about the number of filters used in our convolution operation.

Then the second parameter specifies the size of the convolutional filter in pixels. Filter size may be determined by the CNN architecture you are using – for example, VGGNet exclusively uses (3, 3) filters. If not, use a 5×5 or 7×7 filter to learn larger features and then quickly reduce to 3×3.

The Third parameter specifies how the convolutional filter should step along the x-axis and the y-axis of the source image. In most cases, it's okay to leave the strides parameter with the default (1, 1). However, you may increase it to (2, 2) to reduce the size of the output volume.

The Fourth parameter is the activation parameter which specifies the name of the activation function you want to apply after performing convolution.

5.5 PERFORMANCE ANALYSIS AND EVALUATION

PERFORMANCE ANALYSIS

The performance of face recognition system is analysed based on three parameters.

- A) Training data
- B) Pose variation
- C) Illumination variation in images

TRAINING DATA

The objective of analyzing system by varying data (training-images) is to know the time taken by the system for training the data. This phase includes alignment, feature extraction and training models. Here training is done for 10 persons with 100 images each, 20 persons and so on up to 100 persons with 100 images each. Time is noted while training 1000 images, 2000 images and so on 10,000 images

POSE VARIATION

The performance of the system based on pose variation is done by using dataset of 10 persons. The accuracy is calculated for 10 training images and 50 training images of each person

ILLUMINATION VARIATION

To calculate recognition accuracy for face images showing illumination variations, considered 100 images for training and 100 images for testing. Person name
Recognized faces Accuracy (%)

S.NO	NAME	RECOGNIZED FACES	ACCURACY
1	ACHALA	10	100
2	HARIKA	10	100
3	MAJID	10	100
4	PRASANNA	6	60
5	PRATAP	10	100
6	PRAVEEN	9	90
7	RAKESH	6	60
8	RANADHEER	10	100
9	SHARATH	10	100

PERFORMANCE EVALUATION

The model gives 65-66% accuracy on validation set while training the model. The CNN model learns the representation features of emotions from the training images. Below are few epochs of training process with batch size of 64. Recognition of faces relies on how flexible the system is for pose variations. If the aim of the system is to recognize only frontal faces, then just use few classifiers and function. The number of images of each face relies on the training on an image and testing on the rest. The recognition from different angles depends on the type of images and training set accordingly, with at least one image for each pose per person. The number of images in the training set and test on the remainder depends on the application of the system

Database	Images	Accuracy
Caltech faces	450	85
The CMU database	750000	78
NIST database	3248	96

Seminal contributions were made in¹¹, which discussed the multi-class classification technique as a way to improve accuracy. At the outset, the researchers extracted multiple features from the time-domain of EEG signals, the frequency-domain, and time-frequency features. In turn, the researchers used a variant of the particle swarm optimization (PSO) algorithm for feature selection. This algorithm was improved using a multi-stage linearly decreasing inertia weight (MLDW). The researchers used SVM to classify four emotional states: high-arousal-high-valence, high-arousal-low-valence, low-arousal-high-valence, and low-arousal-low-valence. The average accuracy was 76.67%.

Liu et al., proposed a deep generalized canonical correlation analysis with an attention mechanism (DGCCA-AM), with emotion-related attention mechanism for feature fusion, to five-category emotion recognition¹². A differential entropy (DE) features are extracted from EEG signals using a short-time Fourier transform (STFT) with 4 s nonoverlapping Hanning window. The proposed method was evaluated on a public multimodal dataset, SEED-V. The experimental results reached an accuracy of 82.11% for five emotion classifications with three modalities.

RESULT AND CONCLUSION

Facial expression plays an important role in communication and thus identifying the correct expression is as essential as knowing the exact matter of the communication. This project proposes an approach for recognizing the category of facial expressions. Face Detection and Extraction of expressions from facial images has been achieved successfully and is useful in many applications, such as robotics vision, video surveillance, digital cameras, security and human-computer interaction. This project's objective is to develop a FACE RECOGNITION AND EMOTION RECOGNITION implementing the computer visions and enhancing the advanced feature extraction and classification in face expression recognition. To analyze the performance of the algorithm, extended Cohn-Kanade expression dataset was used initially. Dataset had only 486 sequences with 97 posers, causing accuracy to reach up to 45% maximum. To overcome the problem of low efficiency, multiple datasets were downloaded from the Internet [32, 33], and also author's own pictures at different expressions were included. As the number of images in dataset increases, the accuracy also increased. We kept 70% of 10K dataset images as training and 30% dataset images as testing images. In all 25 iterations were carried out, with the different sets of 70% training data each time. Finally, the error bar was computed as the standard deviation. Figure 4a shows the optimization of the number of layers for CNN. For simplicity, we kept the number of layers and the number of filters, for background removal CNN (first-part CNN) as well as face feature extraction CNN (the second-part CNN) to be the same. In this study, we varied the number of layers from 1 to 8. We found out that maximum accuracy was obtained around 4. It was not giving good accuracy. Hence, FEREC was designed with four layers and four filters. As a future scope of this study, researchers can try varying the number of layers for both CNN independently. Also, the vast amount of work can be done if each layer is fed with a different number of filters. This could be automated using servers. Due to computational power limitation of the author, we did not carry out this study, but it will be highly appreciated if other researchers come out with a better number than 4 (layers), 4 (filters) and increase the accuracy beyond 96%, which we could achieve. Figure 4c and e shows regular front-facing cases with angry and surprise emotions, and the algorithm could easily detect them. The only challenging part in these images was skin tone detection, because of the grayscale nature of these images. With color

images, background removal with the help of skin tone detection was straightforward, but with grayscale images, we observed false face detection in many cases. Image, as shown in Fig. 4g, was challenging because of the orientation. Fortunately, with 24 dimensions EV feature vector, we could correctly classify 30° oriented faces using FERC. We do accept the method has some limitations such as high computing power during CNN tuning, and also, facial hair causes a lot of issues. But other than these problems, the accuracy of our algorithm is very high (i.e., 96%), which is very intuitive, as we assume the number of layers is directly proportional to accuracy and inversely proportional to execution time. Hence due to maximum accuracy obtained with 4 layers, we selected the number of layers to be 4. The execution time was increasing with the number of layers, and it was not adding significant value to our study, hence not reported in the current manuscript. Figure 4b shows the number of filters optimization for both layers. Again, 1–8 filters were tried for each of the four-layer CNN networks. Seven different facial expressions of different persons' images from different datasets have been analyzed. This project involves facial expression pre-processing of captured facial images followed by feature extraction. Local Binary Patterns and classification of facial expressions based on training of datasets of facial images based on Support Vector Machines. This project recognizes more facial expressions based on Haarcascade and fer2013 face. Eyeris is an emotion recognition company that works with embedded system manufacturers including car makers and social robotic companies on integrating its face analytics and emotion recognition software; as well as with video content creators to help them measure the perceived effectiveness of their short and long form video creative. Many products also exist to aggregate information from emotions communicated online, including via "like" button presses and via counts of positive and negative phrases in text and affect recognition is increasingly used in some kinds of games and virtual reality, both for educational purposes and to give players more natural control over their social avatars. Statistical methods commonly involve the use of different supervised machine learning algorithms in which a large set of annotated data is fed into the algorithms for the system to learn and predict the appropriate emotion types. Machine learning algorithms generally provide more reasonable classification accuracy compared to other approaches, but one of the challenges in achieving good results in the classification process, is the need to have a sufficiently large training set.

Conclusion?

REFERENCES: -

1. Black, M. J. and Yacoob, Y. Tracking and recognizing rigid and non-rigid facial motions using local parametric model of image motion. In Proceedings of the International Conference on Computer Vision, pages 374-381. IEEE Computer Society, Cambridge, MA, 1995.
2. Boersma, P., Weenink, D., Praat Speech Processing Software, Institute of Phonetics Sciences of the University of Amsterdam. <http://www.praat.org>
3. Burges, C. A tutorial on support vector machines for pattern recognition. *Data Mining and Know. Disc.*, vol. 2(2), pp. 1--47, 1998.
4. Chen, L.S., Huang, T. S., Miyasato T., and Nakatsu R. Multimodal human emotion / expression recognition, in *Proc. of Int. Conf. on Automatic Face and Gesture Recognition*, (Nara, Japan), IEEE Computer Soc., April 1998
5. Chen, L.S., Huang, T.S. Emotional expressions in audiovisual human computer interaction. *Multimedia and Expo, 2000. ICME 2000. 2000 IEEE International Conference on*, Volume: 1, 30 July-2 Aug. 2000. Pages: 423 -- 426
vol.1
6. Cowie, R., Douglas-Cowie, E., Tsapatsoulis, N., Votsis, G., Kollias, S., Fellenz, W., Taylor, J.G. Emotion recognition in human-computer interaction. *Signal Processing Magazine, IEEE*, Volume: 18, Issue: 1, Jan 2001. Pages: 32 -- 80
7. De Silva, L. C., Miyasato, T., and Nakatsu, R. Facial Emotion Recognition Using Multimodal Information. In *Proc. IEEE Int. Conf. on Information, Communications and Signal Processing (ICICS'97)*, Singapore, pp. 397--401, Sept. 1997.

8. De Silva, L.C., Ng, P. C. Bimodal emotion recognition. Automatic Face and Gesture Recognition, 2000. Proceedings. Fourth IEEE International Conference on, 28-30 March 2000. Pages: 332 -- 335.
9. Dellaert, F., Polzin, T., Waibel, A. Recognizing emotion in speech. Spoken Language, 1996. ICSLP 96. Proceedings. Fourth International Conference on, Volume: 3, 3-6 Oct. 1996. Pages: 1970 -- 1973 vol.3.
10. Ekman, P., Friesen, W. V. Facial Action Coding System: A Technique for Measurement of Facial Movement. Consulting Psychologists Press Palo Alto, California, 1978.
11. Essa, Pentland, A. P. Coding, analysis, interpretation, and recognition of facial expressions. IEEE Transc. On Pattern Analysis and Machine Intelligence, 19(7):757--763, JULY 1997.
12. Huang, T. S., Chen, L. S., Tao, H., Miyasato, T., Nakatsu, R. Bimodal Emotion Recognition by Man and Machine. Proceeding of ATR Workshop on Virtual Communication Environments, (Kyoto, Japan), April 1998.
13. Lee C. M., Narayanan, S.S., Pieraccini, R. Classifying emotions in human-machine spoken dialogs. Multimedia and Expo, 2002. ICME '02. Proceedings. 2002 IEEE International. Conference on, Volume: 1, 26-29 Aug.

Facial Expression Recognition via Deep learning

Arunachalam.R

Department of IT

*Sathyabama Institute of Science and
Technology*

Chennai, India

arunachalam2432001@gmail.com

Dr.Mary Gladence M.Tech., Ph.d.,

Department of IT

*Sathyabama Institute of Science and
Technology*

Chennai, India

marygladence.it@sathyabama.ac.in

Ferose Moosha.F

Department of IT

*Sathyabama Institute of Science and
Technology*

Chennai, India

ferosefjm786@gmail.com

ABSTRACT

The human face has idiosyncratic and unique traits, so it becomes tough to perceive and pick out faces. Determining the facial expression of someone in any series of pics is easy. However, if we study the automated recognition gadget, the available systems are completely insufficient and not able to correctly discover movements. The area of face identity has many critical applications. It is an interactive device between people and computer systems. The person can work with his face with out the usage of his hand. Currently, facial studies is targeted on causes which include sadness, pleasure, disgust, marvel, fear, and anger. This article objectives to discover faces in any given picture, extract facial expressions (eyes and lips), and classify them into 6 emotions (happiness, fear, anger, disgust, neutrality, sadness). The education data goes thru a chain of threads and techniques and is sooner or later characterized through a vector machine (SVM) subtle by means of research. The check records then exams the statistics and labels and offers an accurate type of the take a look at information within the category report. Various approaches had been applied, together with passing fashioned pics via a Gabor clear out or reworking snap shots the usage of a histogram of degrees (HOG) and discrete wavelets into better information class (DWT). The high-quality outcomes to date had been performed by means of running photo workouts thru Histogram of Eastern Gradients (HOG) characterization followed with the aid of SVM, supplying a median accuracy of 85%.

Keywords:

Facial Expression Recognition, Deep Learning

INTRODUCTION

In the procedure of human-computer interplay, instructors evidently count on that the face reputation system can come across, examine and process emotions to reap a good studying impact including perception, expertise and expression of emotions. By showing unique facial expressions within the video, teachers can decide if college students apprehend the mastering fabric thru extraordinary student expressions and may adjust their curriculum.

In a digital environment, information on pupil motion can be acquired from a chain of video frames. Therefore, if you want to obtain a real-time expertise of students' emotions, it's miles important to balance accuracy and time. To improve the accuracy, more video frames are required, which will increase the computation time. On the other hand, in the pursuit of performance, we need to lessen accuracy, which means gathering less unique statistics.

LITERATURE SURVEY

INFERENCES FROM LITREATURE SURVEY

This method of the trouble has been cautiously studied for five years by researchers and car organizations to find a answer, and all their answers variety from analyzing various sorts of distraction to analyzing the driver's vital signs and symptoms

Capture-the-invisible (CTI): Behaviour-based totally detection in IoT-centric business manipulate structures.

Author: Akashdeep Bhardwaj, Fadi Al-Turjman

Year: 2020

Description: Industrial manage structures display, automate and manage complex infrastructure and techniques which might be incorporated into crucial industries that have an effect on our each day lives. With the appearance of networking and automation, these systems have evolved from committed and impartial centralized infrastructure initiatives. While this tracking and average control is easier the usage of conventional detection strategies, internet application firewalls or intrusion detection systems have made exposed networks vulnerable to cyberattacks. Such assaults alternate the go with the flow of manipulate and processes and feature the malicious capacity to absolutely alternate those systems. This look at makes a speciality of the use of analytical tactics for assault detection in commercial control infrastructure systems and compares assault-based totally detection methods. The proposed paintings version is an invisibility reputation algorithm referred to as Capture of the Invisible.

(CTI) "to discover hidden processes in the bushes of commercial manipulate systems and to discover conduct-based attacks in actual time.

Image Forensics Ex luce volutpat

Author: Manoj Kumar, Anuj Rani and Sangeet Srivastava.

Year: 2019

Description: Computer-generated pictures are considered to play a key position in anyone's lifestyles in this age of information generation, as human beings efficaciously attend advertising, magazines, websites, television and greater. Now that virtual photos have performed their function, the probabilities of slander, abuse, misuse are blocked and additionally come to be less complicated via picture editing programs. To be criminal, if someone has carried out some thing incorrect the proposed approach may be used to properly discover fakes and impersonate them in digital photographs. In existing strategies, researchers have proposed the most commonplace methods of digital photograph manipulation consistent with the supply, metadata, picture description, gluing and greater. The proposed technique is based totally on physical methods and requires less human involvement.

Secure Attribute-Based Communication for Resource Constrained Users in Cloud Computing

Author: JinLia, YinghuiZhangbcd

Year: 2018

Description: Data alternate is turning into a totally attractive provider thru cloud computing because of its convenience and fee-effectiveness. Attribute-based encryption (ABE) has obtained huge attention as a potential technique for imposing great-grained communication. However, maximum of the existing ABE solutions be afflicted by the risks of excessive computational fees and susceptible facts protection, which significantly prevent resource-restrained mobile gadgets from configuring the provider. The trouble of simultaneously reaching granularity, high efficiency on the a part of the facts proprietor and records privacy standards within the change of cloud records is still doubtful. This paper solves this difficult problem by using proposing a new characteristic-based totally verbal exchange version suitable for mobile users in cloud computing. The proposed solution removes a lot of the computational burden via adding public device parameters as well as taking partial encryption calculations offline. In addition, a public ciphertext verification section is achieved earlier than the decryption phase, which removes a good deal of the computational overhead because of illegitimate signatures. For information security, the Chamaeleon Hash feature is used to mechanically generate the ciphertext in order to be used to produce the last acknowledged data from the offline statistics. The proposed scheme has been established to be comfortable against the assaults of adaptive choice ciphertext, that's broadly general as a preferred safety concept. An extended exposure analysis shows that the proposed software is safe and powerful.

Achieving Attribute Confidentiality and Fast Decryption for Data Security in Mobile Cloud Computing

Author: Yinghui Zhang, Xiaofeng Chen

Year: 2017

Description. Although many users are transferring their diverse data to the cloud, records protection and privateness problems are nonetheless the largest barriers stopping the sizeable adoption of cloud computing. Anonymous Attribute-Based Encryption (Anonymous ABE) gives unambiguous get admission to to manage to cloud garage and continues the recipient's attributes private via hiding attribute records in ciphertexts. However, within the present anonymous ABE operation, the person will simplest recognize if the

attributes and the hidden coverage match or no longer after a minute. In addition, many pairs are typically required for each decryption, and the computational value will increase with the complexity of the access method. Existing schemes be afflicted by a serious lack of efficiency and are not suitable for mobile cloud computing, where users can be confined in assets. The decryption level. This method evaluates the unique functions within the characters used to test if the personal key attribute complies with the privacy coverage within the characters with out decryption. For rapid decryption, unique characteristic secret keys are generated that permit pairs to be aggregated in decryption. We provide a simple anonymous ABE construct after which take a extra cozy extension from one-time non-existant signatures. In the proposed scheme, the computational cost of matching characteristic evidence is much less than a unmarried decryption operation that calls for a small and consistent range of pairs. Formal safety analysis and performance comparisons show that the proposed solutions both make certain privacy and improve the effectiveness of decryption for external information stored in mobile cloud computing.

Title 5: Divorce

Author: ZheliLiua, YanyuHuang

Year: 2018

Description. Oblivion RAM (ORAM) is vital for packages that require hidden access styles. Many ORAMs had been proposed, however most handiest guide blocks of the same length. Because of the various period of the information blocks, they are commonly packed with only that length before receiving a discount, ensuing in more garage area and community bandwidth usage. In order to increase the first sensible block of variable size ORAM, we proposed "DivORAM", building on the ORAM shape. It implements an additive homomorphic encryption scheme (Damgord-Eurek cryptosystem) at the server aspect to reduce client computing fees and community flexibility. As a result, the bandwidth of the community is reduced by means of 30% as compared to Ringone ORAM and 40% to HIRB ORAM. Experimental outcomes show that DivORAM reaction time is better than Ring ORAM for sensible modules.

OPEN PROBLEMS IN EXISTING SYSTEM

ILLUMINATION

A small trade in lighting fixtures situations creates a critical trouble for face recognition itself and might notably affect its consequences. If the sitting has a

tendency to change, the same person stuck by using the equal feeling with a face and face and conduct almost comparable, the end result can be absolutely one-of-a-kind.

POSE

Facial recognition systems are very sensitive to modifications in posture. The attitude of the face changes with the motion of the head and the angle of the man or woman's gaze. Head movements or specific camera angles can irreversibly cause modifications inside the facial expression and generate versions in the elegance, resulting in a drastic drop in the pace of facial reputation. As the attitude of rotation increases, it's miles hard to understand a real face. This can cause know-how or no understanding if best a face view is available in the database.

RESOLUTION

The minimal decision of any preferred photo ought to be 16*16. A image with a resolution smaller than sixteen*sixteen is referred to as a low resolution photo. These low-decision photos may be captured using small stand-alone cameras which includes road surveillance cameras, device cameras, and grocery store security cameras. These cameras can only seize a small part of a person's face, and because the digital camera is not close to the face, they could only capture regions of the face smaller than sixteen*16. The photograph of this kind of low spirit does no longer deliver many clues that the general public are misplaced. This may be a massive problem inside the face popularity process.

AGEING

The appearance/texture of the face adjustments through the years and ages, that's another trouble in facial recognition. With age, human facial features, shapes/strains and other factors also alternate. This is finished for visible remark and image retrieval after a long time.

To take a look at the accuracy, the dataset is calculated for a one of a kind age institution of humans over a positive time period. This recognition system relies upon on the extraction of capabilities, painful capabilities which includes wrinkles, marks, eyebrows, hair, and so on.

PRIVACY

Privacy worries exist with facial popularity technology due to its capability to pick out and get dressed a place that can constitute an intrusion. Facial popularity era can tune someone, which makes human beings aggravating. Also, extreme data breaches are all too commonplace nowadays and personal statistics accrued via facial reputation software program isn't cozy.

FACIAL EXPRESSION DETECTION VIA DEEP LEARNING

MODULES IN FERDL:

FACE CAPTURING MODULE

At this degree, we seize human beings's faces for similarly processing with a webcam or external picture seize, no processing may be finished and there may be no way to discover movement.

PREPROCESSING MODULE

After shooting the pics, we can manner the pix earlier than. In this step we can convert colour pictures to grayscale photos.

TRAINING MODULE

In this step we are able to put together a dataset that is a binary array of all captured images. The captured pics can be saved in a .YML document, which shops all of the face facts.

FACE RECOGNITION MODULE

The first step inside the face recognition system is the switch of face information to the bottom machine. The laptop gadget's webcam takes 60 distinctive facial pictures. In this module we can perceive the faces the usage of the LBPH algorithm. LBPH is an abbreviation for Local Binary Histogram Model. It recognizes the face with the Face ID and NAME which have been formerly stored.

EXPRESSION RECOGNITION MODULE

Facial recognition software program is a generation that makes use of biometric markers to hit upon actions in people's faces. It extracts and analyzes facts from an photograph and can provide comments or uncorrupted, motion-weighted information.

SOFTWARE REQUIREMENTS IN FERDL

a) IDLE (Python 3.6 64-bit)

b) Java Development Kit

c) OpenCV, Keras, TensorFlow and Tkinter libraries.

DESCRIPTION OF PROPOSED SYSTEM

SELECTED METHODOLOGY OR PROCESS MODEL

FACE DETECTION:

Face detection is an critical FER time. It is feasible that an green computerized device may be developed to apprehend the spontaneous facial area in a static photograph or video. The facial region is detected in a sequence of pictures the use of facial features together with edges, pores and skin color, texture and facial muscle movements.

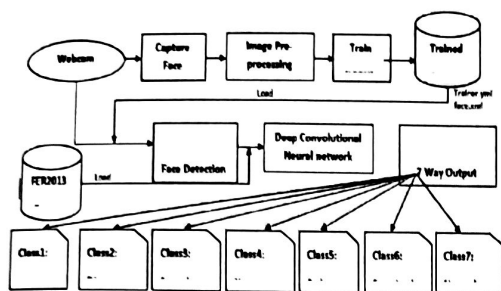
FEATURE EXTRACTION:

After the face popularity, the subsequent step in the FER extraction characteristic. The major motive of facial feature extraction is to acquire an efficient and powerful representation of facial features without dropping information about the face. Geometric functions and face-based features are two kinds of function extraction which might be outstanding via facial motion and facial deformation. The input picture may be either a nonetheless photo or a sequence of snap shots. Based on the input photograph, the proper facial extraction set of rules is implemented to extract either neighborhood, international or hybrid functions. The extracted functions are extensively contracted, which is fed as the input to the classifier, that's most helpful in making the classifier choice less difficult in figuring out and recognizing facial expressions.

EXPRESSION CLASSIFICATION:

Classification is a control model used to are expecting output primarily based on located values. In pattern recognition, the classifier plays a crucial position by means of either inferring or predicting the magnificence labels of unseen information based totally on trained pix. Many class algorithms had been evolved to gain most popularity speed. Indeed, now not all approaches are suitable for the analysis of diverse methods for extracting facial capabilities.

ARCHITECTURE / OVERALL DESIGN OF PROPOSED SYSTEM



FERDL system model architecture

EMOTION RECOGNITION USING CNN

DATA PERPROCESSING

The pixel column is in general stored in list layout. Since the computational complexity of computing pixel values within the range (zero-255) is excessive, the facts inside the pixel discipline is normalized to values between [0-1]. The saved face gadgets are resized and resized to a fixed size of 48 X 48. The corresponding movement labels and their corresponding pixel values are saved inside the gadgets. We use the scikit-study `train_test_split()` function to break up the time table into education and checking out facts. Test_size 0.2 way 20% of the statistics is for trying out and 80% of the information can be exercised.

IMAGE AUGMENTATION

Additional statistics is generated the usage of the formation with the aid of applying ameliorations. The schooling required for schooling isn't always sufficient to teach representation. Image records is generated from the photo itself by using transforming, rotating, cropping, cropping, scaling, flipping, mirroring, normalizing, etc.

FEATURE EXTRACTION

Among all the facial expressions, the localization and detection of the eyes are critical, by means of which the places of all different facial expressions are marked. But the prevailing face popularity methods couldn't apprehend the precise man or woman. To remedy this trouble, Face Recognition Geometric Features Extraction (GSF 2 EFR) is designed to as it should be apprehend a person by using determining the middle and corners of the eye the usage of eye detection and localization modules. An enter face photo is given in which the face map is initialized and runs in hard mode.

DATA TRAINING

2- and 3-dimensional computerized face recognition techniques are finding increasingly more applications. In addition to its original purpose, which is the identity or verification of folks inside the area of security and criminal regulation, many fields of software have been introduced in the financial and studies industries. With facial popularity, robots, for example, can recognize human moves and act therefore. Online eyewear retailers can advise the proper frame for any user thru facial popularity, and clients can strive it on certainly. Through on line automatic face reputation, customized style and makeup hints can be supplied, and a brand new breed may be predicted.

VALIDATION

In the validation segment they used diverse OpenCV features and Keras features. Initially, the video files are saved within the video item. Cascade LBP classifier is used to discover the facial vicinity of hobby. The picture of the body turns grayish, its size and shape alternate. Using numpy. This resized photograph is fed into the predictor, which is loaded with the `keras.Fashions.Load_model()` function. The fundamental argument is the output. A rectangle is drawn around the front areas, and the output is fashioned over a rectangular field.

The convex block is calculated, the triple shaped through the angles on the sides of the two first centroids with the smallest coordinates, and the modulus in their difference is determined. In a success instances, the 2 centroids with the smallest y-coordinates are the 2 eyes. The set of triplets is then used as enter to the kmeans condensing set of rules (with okay = 2) and the containers for each magnificence are extracted, one representing the type set of successful detections and the alternative representing false detections.

FACIAL RECOGNITION WITH MUSIC RECOMMENDATION

Facial and emotion popularity may be completed the use of an expansion of features, which includes facial speech and even textual content. Among these functions, facial expressions are a number of the maximum popular, making it less difficult to attain a huge amount of facial statistics (another tool for human recognition). Recently, with using deep studying and mainly Convolutional Neural Networks (CNs), many

EXPERIMENTAL ANALYSIS AND EVALUATION

EXPERIMENTAL ENVIRONMENT

To explore the ability and effectiveness of numerous Chernovic faces, we conducted a managed experiment with twenty-four topics. The display monitors were created via a Java 1.1 software that displayed Chernoff's face for a sure period of time. Experiments have been performed on a Sun Sparc five.6 pc. This experiment used two designs within topics. The two elements that had been systematically numerous to create the take a look at situations were the objective challenge and the number of faces in the display. Four stages of aces are examined (5, 10, 25 or 50). The order wherein the different face numbers have been supplied became random.

PERFORMANCE ANALYSIS AND EVALUATION

PERFORMANCE ANALYSIS

The overall performance of the face popularity device is based totally on three parameters.

- A) Training data
- B) Pose variation
- C) Illumination variation in images

TRAINING DATA

The purpose of growing a system with distinctive data (host snap shots) is to recognize the time that the machine takes to install the statistics. This consists of a night scene, feature extraction and version training. Here, 10 humans are skilled for one hundred pix, 20 human beings and so forth up to 100 humans for 100 pix. Time is recorded while forming one thousand photos, 2000 photographs, 10,000 pics and so on.

POSE VARIATION

Position-based totally performance analysis is performed the usage of a dataset of 10 individuals. It is carefully calculated for 10 forming snap shots and 50 forming photographs per person.

ILLUMINATION VARIATION

To calculate the accuracy of the recognition of face snap shots showing versions in illumination, one hundred snap shots were taken into consideration for schooling and one hundred photos for testing. Person identified by name Accuracy (%)

S.NO	NAME	RECOGNIZED FACES	ACCURACY
1	ACHALA	10	100
2	HARIKA	10	100
3	MAJID	10	100
4	PRASANNA	6	60
5	PRATAP	10	100
6	PRAVEEN	9	90
7	RAKESH	6	60
8	RANADHEER	10	100
9	SHARATH	10	100
10	SRINU	10	100

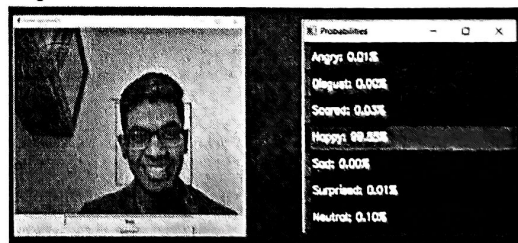
PERFORMANCE EVALUATION

The version offers sixty five-sixty six% accuracy in validation set whilst the model is installed. The CNN model studies the features of representing emotions from habitual pictures. Below are numerous schooling system epochs with a batch length of 64.

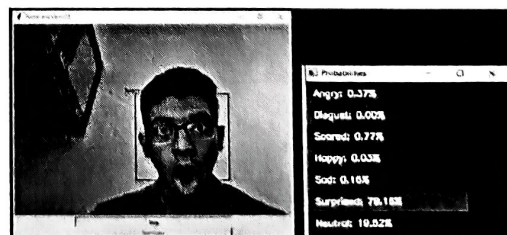
Facial reputation relies upon on how bendy the system is for posture versions. If the intention of the device is to understand faces only from the front, just use a few classifiers and features. The number of snap shots of every face depends on the photo and the experiment on the rest. The recognition from distinct angles relies upon on the forms of photographs and the installation, respectively, at the least one image for each pose in step with man or woman. The number of files covered within the installation and the residual take a look at relies upon on the software region of the system.

RESULT AND CONCLUSION

Facial expression performs an essential function in conversation, so accurate expression identity is as critical as understanding the exact essence of the message. This undertaking proposes a class recognition approach for facial expressions. Face detection and expression extraction from facial photographs have been efficaciously implemented and used in lots of programs inclusive of robotic imaginative and prescient, surveillance imaginative and prescient, virtual cameras, protection, and human-computer interaction. The purpose of this challenge is to broaden FACE RECOGNITION AND MOVEMENT RECOGNITION a good way to improve laptop imaginative and prescient implementation and feature extension.



And department in face recognition.



Seven facial expressions of various pics of different human beings from exceptional datasets were analyzed. This method includes pre-processed taking pictures the expressions of the photographs observed through the characteristic extraction. Spatial binary styles and facial voice classification primarily based on training face

photograph datasets based on assist vector machines. This mission recognizes numerous facial expressions in Haarcascade and fer2013 faces.

REFERENCES:-

- 1) Black, M. J. and Yacoob, Y. Tracking and recognizing rigid and non-rigid facial motions using local parametric model of image motion. In Proceedings of the International Conference on Computer Vision, pages 374-381. IEEE Computer Society, Cambridge, MA, 1995.
- 2) Boersma, P., Weenink, D., Praat Speech Processing Software, Institute of Phonetics Sciences of the University of Amsterdam. <http://www.praat.org>
- 3) Burges, C. A tutorial on support vector machines for pattern recognition. *Data Mining and Know. Disc.*, vol. 2(2), pp. 1--47, 1998.
- 4) Chen, L.S., Huang, T. S., Miyasato T., and Nakatsu R. Multimodal human emotion / expression recognition, in *Proc. of Int. Conf. on Automatic Face and Gesture Recognition*, (Nara, Japan), IEEE Computer Soc., April 1998
- 5) Chen, L.S., Huang, T.S. Emotional expressions in audiovisual human computer interaction. *Multimedia and Expo, 2000. ICME 2000. 2000 IEEE International Conference on*, Volume: 1, 30 July-2 Aug. 2000. Pages: 423 -- 426 vol.1
- 6) Cowie, R., Douglas-Cowie, E., Tsapatsoulis, N., Votsis, G., Kollias, S., Fellenz, W., Taylor, J.G. Emotion recognition in human-computer interaction. *Signal Processing Magazine, IEEE*, Volume: 18, Issue: 1, Jan 2001. Pages: 32 -- 80
- 7) De Silva, L. C., Miyasato, T., and Nakatsu, R. Facial Emotion Recognition Using Multimodal Information. In *Proc. IEEE Int. Conf. on Information, Communications and Signal Processing (ICICS'97)*, Singapore, pp. 397--401, Sept. 1997.
- 8) De Silva, L.C., Ng, P. C. Bimodal emotion recognition. *Automatic Face and Gesture Recognition, 2000. Proceedings. Fourth IEEE International Conference on*, 28-30 March 2000. Pages: 332 -- 335.
- 9) Dellaert, F., Polzin, T., Waibel, A. Recognizing emotion in speech. *Spoken Language, 1996. ICSLP 96. Proceedings. Fourth International Conference on*, Volume: 3, 3-6 Oct. 1996. Pages: 1970 -- 1973 vol.3.
- 10) Ekman, P., Friesen, W. V. *Facial Action Coding System: A Technique for Measurement of Facial Movement*. Consulting Psychologists Press Palo Alto, California, 1978.
- 11) Essa, Pentland, A. P. Coding, analysis, interpretation, and recognition of facial expressions. *IEEE Transc. On Pattern Analysis and Machine Intelligence*, 19(7):757--763, JULY 1997.
- 12) Huang, T. S., Chen, L. S., Tao, H., Miyasato, T., Nakatsu, R. Bimodal Emotion Recognition by Man and Machine. *Proceeding of ATR Workshop on Virtual Communication Environments*, (Kyoto, Japan), April 1998.
- 13) Lee C. M., Narayanan, S.S., Pieraccini, R. Classifying emotions in human-machine spoken dialogs. *Multimedia and Expo, 2002. ICME '02. Proceedings. 2002 IEEE International Conference on*, Volume: 1, 26-29 Aug.