Three-Dimensional Emoji Prediction Using Facial Emotion Recognition

T.U. Dalugama
2019

Three-Dimensional Emoji Prediction Using Facial Emotion Recognition

A dissertation submitted for the Degree of Master of Science in Computer Science

T.U. Dalugama
University of Colombo School of Computing
2019

Declaration

The thesis is my original work and has not been submitted previously for a degree at this or any other university/institute.
To the best of my knowledge it does not contain any material published or written by another person, except as acknowledged in the text.

Student Name:	T. U. Dalugama
Registration Number:	2016/MCS/018
Index Number:	16440181

Signature: Date:

This is to certify that this thesis is based on the work of
Ms. T.U. Dalugama

under my supervision. The thesis has been prepared according to the format stipulated and is of acceptable standard.

Certified by:
Supervisor Name: Dr. G.D.S.P. Wimalaratne

Signature: Date:

Abstract

Reliable communication relies on thoughts of human while communicating with a group of people. Facebook, Instagram, and Twitter are some of the social media where various types of human emotions are being shared. Two-dimensional emojis are widely used in communication rather than sharing human faces on social media. Sharing two-dimensional emojis is more comfortable without sharing live human facial emotions using a video stream.

As people do not share their own real-time videos on social media, it is more realistic if people can share three-dimensional emojis based on real-time facial emotions on social media. Real-time facial emotion recognition using facial expressions helps to embed three-dimensional emojis with relevant facial expression type. This thesis investigates human facial emotion recognition using facial expressions and integrating real-time facial emotions with three-dimensional emojis.

Modules of the research design consists of face detection module, emotion classification module and three-dimensional emoji prediction module. Face detection module detects the face of a human using a real-time video stream. Emotion classification module uses the detected human face to extract the human facial expression with facial emotions. Three-dimensional emoji prediction module consumes extracted facial emotion type to predict associated three-dimensional emoji to the user.

Evaluation of the research study was conducted using a sample group. The response given by the sample group was used to analyze the research with results. Evaluation process was performed via a selected set of parameters gained by the research methodologies. The accuracy level and effectiveness level of the emotional change on 3D emoji were calculated to evaluate the research study to maintain a better relationship with the input of the research study and the expected output.

This research focused on three-dimensional emoji prediction using facial emotion recognition. Research study emphasized on improving human machine interaction between social communication media.

Acknowledgments

This thesis is the result of me being fortunate to have the assistance of several people who have been extremely supportive in various ways. First and foremost, I would like to offer my humble gratitude to Dr. Prasad Wimalarthne, my supervisor, for being extremely supportive and for the guidance he had given me. He had been a great listener, a great advisor and a great teacher.

I wish to thank all the lectures at the University of Colombo School of Computing for their valuable advices and comments. I would like to thank my beloved parents and my husband for guiding me with the love and affection to fulfil my research. Finally, I would like to thank all the people who have helped me throughout this project.

Table of Content

List of Figures

List of Tables

Chapter 1

Introduction

Usage of emojis are a trend of all types of social media communications. Nor Facebook, Twitter, Instagram can survive without emojis. Emojis are important in day to day conversations [27], which take place in our daily lives. Because of high usage of emojis people have declared a world emoji day to denote the importance of emojis. Commonly used emojis are two-dimensional emojis. Three-dimensional animations can effectively give a more realistic feeling rather than two-dimensional emojis.

Three-dimensional emojis can give a more realistic experience to the user [37]. Numerous researches have been conducted by Massachusetts Institute of Technology (MIT) on social media and emojis [24]. This research attempt focuses on three-dimensional emojis with emotion variations [38]. Emotions of three-dimensional emojis vary based on real-time human facial expressions and human facial emotions.

1.1 Problem definition

The interaction between human beings and mobile devices will be more natural if mobile devices are capable to perceive and respond to human non-verbal communication [5] such as emotions. People do not like applications with fewer interactions [38]. The reason is that human beings are more interested in realistic feeling while communication via mobile devices or computer applications [28]. The realistic feeling will give the satisfaction of sharing their thoughts, feelings, and emotions as they are doing text message communication [5]. Users like realistic feeling in computer applications. Facial emotions are very tenuous features in a human face [3]. Therefore applications that identify happiness, sadness, disgust, surprise, anger, and fear [1] of the face and react according to the facial expression will give users a sensible and realistic feeling [2].

However, with the rise of mobile devices and chat applications, emojis have become an increasingly important part of written language a well know an effective way of communication. Text messaging and chat applications are ways of communication mechanisms in the modern world. People used to communicate with one person or a group of people.

People are used to having text chats with other parties rather than having video calls while working at the office or traveling. Emoji icons are more useful when people are having chats with others. Image icons and emoji icons are used to convey the feelings and emotions with each other as the text cannot express the current feelings and emotions as it is [25]. Therefore the human-computer interaction is more important in communication [23].

1.2 Background

Most of the emoji icons are designed based on human facial expressions and emotions. Therefore most of the mobile device users and computer application users are using emoji icon images on the SMS communication. Most of the social media applications have recognized that the usage of emoji icons has been increased over time. Therefore social media like Facebook, Twitter and Instagram have declared world emoji day in order to celebrate the usage of emojis [26]. Communication with smiles is more powerful than only sending text messages while on communication.

The world is more focused on three-dimensional objects rather than two-dimensional objects. This is because of the realistic feeling that is being given by three-dimensional objects to the users [24]. Therefore three dimensional (3D) emojis gives a closer relationship of emotional communication rather than using two dimensional (2D) emoji icons [38].

The research is based on facial emotion detection and applies emotional 3D emoji for the corresponding facial emotions based on facial expressions. Graphical representations embedded in online communication can effectively convey information without relying excessively on the text. Therefore 3D emoji and graphical representation combined with facial emotion will improve the online communications among people [5]. The research will be conducted on emotional 3D emoji applications based on relevant facial emotions of human beings and the graphical representation of online communication [4].

Identification of facial emotions such as happiness, sadness, disgust, surprise, anger and fear of the human image using a live video stream and adopt 3D emoji according to the detected facial emotion from the video stream will provide realistic feeling to the users. 3D emoji adoption is basically based on the graphically changing environment of the emotions in order to express the real-time feeling of users on communication. Facial feature extraction methodologies will be used to classify facial emotions using facial expressions. Within the scope of the project, emotions will be identified using feature extractions. Real-time graphical changes to the emoji help for interactive communication between users.

3D emoji can be used in communication media with graphical changing ability embedded in online communication [5] with 3D emoji. This helps in effectively conveying information without relying excessively on the text. Therefore real-time interaction is more important to give a realistic feeling to users [22]. Contribution mainly focuses on interactive communication and real-time interaction with other parties based on emotion variations.

This provides the path to have effective communication with people who are not closer to users. Study the application of Cloud vision API for emotion recognition of humans using facial expressions. Recognize 3D emoji applications and variation of emoji expressions that correspond with human facial emotions. Facial emotions and 3D emoji applications are useful for human-computer interaction [23]. Extraction of facial features using eyebrows, eyes, teeth, lips [21], directions of specific facial muscles to identify facial emotions of human face [3].

1.3 Aims and objectives

The aim of this research is to exchange emotionally embedded 3D emojis on a user's emotional variations for efficient communication. To achieve the goal, this research investigates and develops a methodology to animate three-dimensional emoji with real-time human facial expressions and emotions. Human facial emotions vary within a very short time period. These facial emotions and facial expressions can be embedded in three-dimensional emojis in communication [40], where people can share real-time emotions with each other. To achieve this aim there are some objectives to accomplish.

- To investigate facial expression using face detection approaches.
- To review the literature on human-machine interactions using three-dimensional emoji.
- To enhance a three-dimensional emoji model to vary with facial features.
- To conduct a usability evaluation experiment on the data analysis

1.4 Research questions

1. What are the ways of identifying the emotions of the users from frequent frames of real-time video stream [1], [17], [2]?
2. How to categorize emotions of 3D emoji based on emotion types of the human face [4], [5], [19]?

There are multiple ways of techniques to detect human face. The focus is to find a more accurate method to detect the face and emotions to predict the most relevant three-dimensional emoji.

1.4 Scope

The research is about three-dimensional emoji prediction using facial emotion recognition. The method of Facial feature classification is used to identify face and facial emotions. Classification algorithms are used to recognize facial emotions types based on human facial variations [1]. The method of facial emotion recognition (FER) [18] can be used to compile detected faces using input images from video clips and classify facial expressions from feature extraction [17]. Emoji is used to express emotional feelings to other parties. Therefore emoji classification will be conducted to investigate emotional recognition using a facial expression by emoji [4]. Real-time interactive communication between gives users a realistic feeling. As adhere to the methodologies graphical changing ability of 3D emoji should be adopted to complete research on real-time graphical changes on 3D emoji.

1.5 Organization of the thesis

Chapter two reviews the other related works to this project and the background of the project. Chapter three discusses the design architecture of the project in detail. Chapter four presents the implementation of the research. Chapter five illustrates the evaluation results and chapter six concludes the dissertation with a conclusion and a discussion about future work.

Chapter 2

Literature review and background

2.1. Introduction

This chapter explains existing approaches that can be considered to the proposed approach. This chapter provides a brief description of how existing approaches are differing from the solution. Basically, the chapter focuses on the main features of the solution compared with other existing approaches. Further, this chapter provides a justification of the existing technologies and critical review of each technology type.

2.2. Related work

There are many technologies that were used for face detection and facial emotion recognition. Related work explains each technology and the comparison among technological approaches.

2.2.1. Google cloud vision API

Google cloud vision API [14] is an artificial intelligent and machine learning related product. Google cloud vision API can be used to derive insight from an image. There are pre-trained API models to detect a broad range of objects in an image. Google cloud vision API [49], has embedded easily train custom vision models with AutoML Vision.

The content of an image is described to the developer by encapsulating powerful machine learning models. There is an easy-to-use REST API embedded in Google cloud vision API [42]. The classification techniques are used to detect faces and objects inside a wide range of photograph that contains lots of objects inside.

Google cloud vision API is enabled of reading printed words within the image. The accuracy level is high in vision API. The AutoML Vision [14] is capable of grabbing concepts from the images. Similar concepts can be categorized into a type of categories using vision API.

As shown in figure 2.1: Facial emotion recognition via cloud vision API, the drag and drop section of the Cloud vision API can detect emotions of humans by analyzing the dropped image into the API. Even Figure 2.1: Facial emotion recognition via cloud vision API, contains multiple objects it is capable to identify each object in the image and analyze the facial expression of the image at the same time.

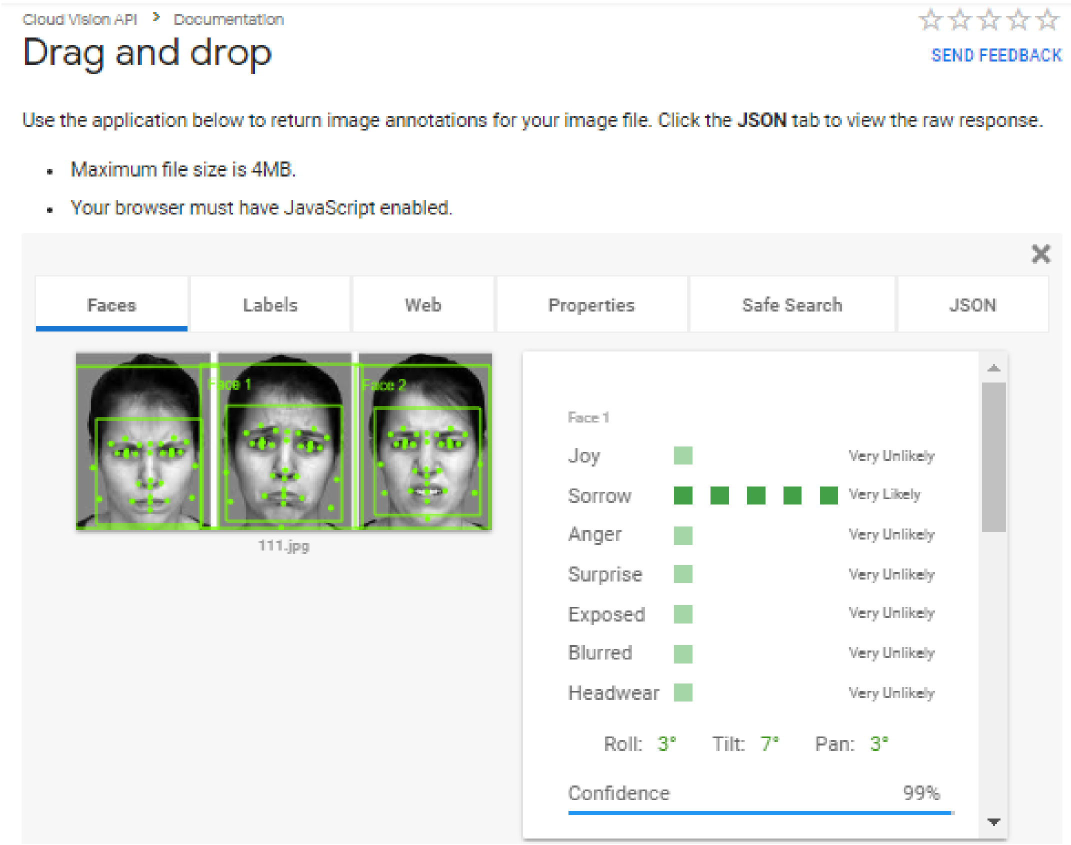


Figure 2.1: Facial emotion recognition via cloud vision API (Yik, Widen, & Russell, 2013)

There are three faces in the image. As shown in Figure 2.1: Facial emotion recognition via cloud vision API, the Cloud vision API has identified three faces as two faces and the first two faces are described as under the face category. The first two images were not described under separate descriptions. The third image is described under the second face section.

The JSON output of the analyzed image implies the view of the raw response [15]. All the related data is displayed in the JSON format as raw data. Face Detection in Google cloud vision API detects multiple faces within an image. Google cloud vision API is capable of providing coordinates of key facial features. This feature describes and predicts the emotion [33] state of the face inside the image. The emotion type can be joy, anger, and surprise. FACE_DETECTION is the feature type of face detection.

Figure 2.2 shows the most related emotion category for the image. The relatedness of the category depends on the levels such as very unlikely, unlikely, likely and very likely. Google vision API is capable of categorizing into emotion groups and gives the tendency for each emotion category when detailing the emotions of the image [50]. The emotion categories are joy, sorrow, anger, and surprise. The emotion of the human portrait is identified based on the feature point of the face. The emotion defers with the eyes, mouth and other facial muscle locations.

The relative emotion category of the image is useful to predict the particular matching emoji based on the interactive user's facial expressions [30]. The 3D emoji can be predicted based on the emotion type of the user to improve the human-computer interaction [33].

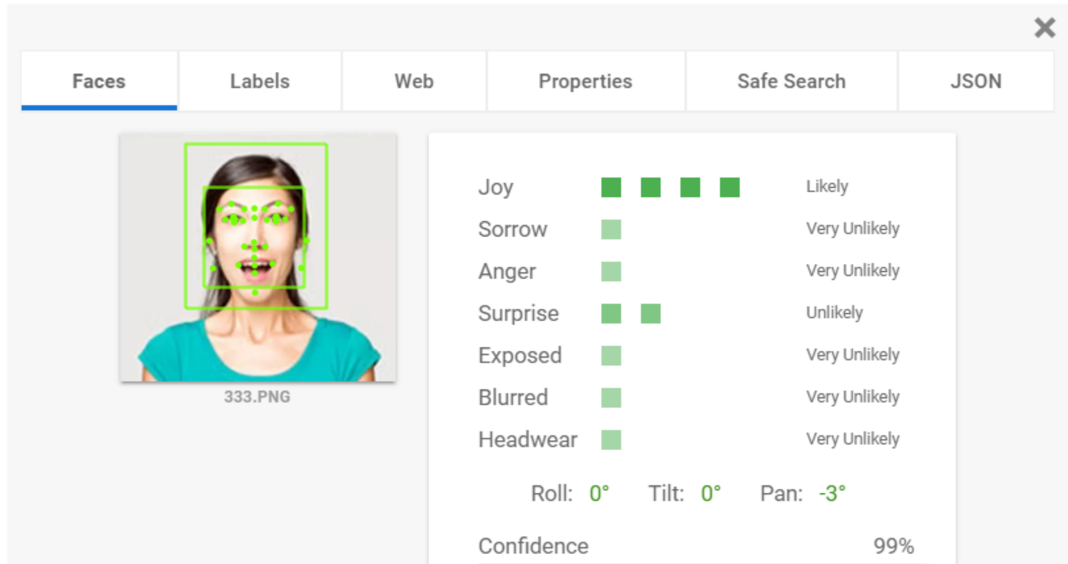


Figure 2.2: Facial emotion type recognition via cloud vision API (Titcomb, 2016)

As shown in Figure 2.2: Facial emotion type recognition via cloud vision API, the feature points are used to detect eyes, nose, mouth, and eyebrows of the portrait of the image [51]. The image can extract from a live video stream to detect the facial emotion recognition using facial expressions.

2.2.2. Deep emoji

Deep emoji research was conducted by the Massachusetts Institute of Technology (MIT) [16]. As emotion content is an important part of the language it is great if emotion can be detected and predicted based on the sentence. The detection of emotion is conducted by machine learning techniques. Machine learning is asking the machine to learn from many examples without explicitly stating the output.

MIT has trained its model to predict emojis on a dataset of 1.2 billion tweets [16]. Deep emoji related research used neural networks which are based on artificial intelligence techniques [35]. The artificial intelligence was focused to rate the emotion of the human as they writing on the tweet.

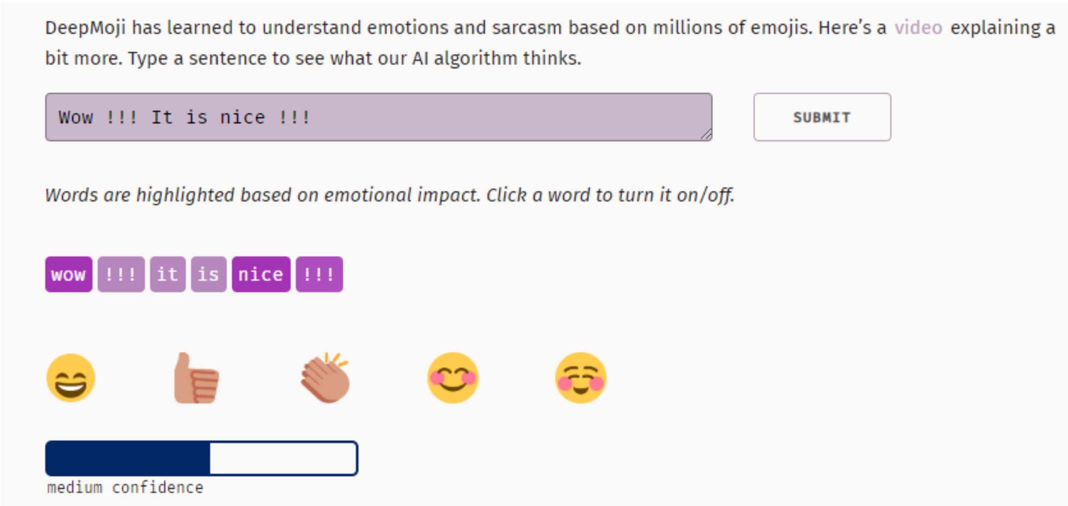


Figure 2.3: Emoji prediction on happy sentences

As illustrated in Figure 2.3: Emoji prediction on happy sentences and Figure 2.4: Emoji prediction on unhappy sentences [44], the user can type sentences and submit them. All the related and predicted emojis appear below entered text sentences [33]. It is hard to predict emojis based on the typed sentences by a human. MIT has provided a word emphasis section above the predicted emojis [34]. The highlighted word defines that used to predict particular emojis. Artificial intelligence was used to learn about people's emotions through tweets.

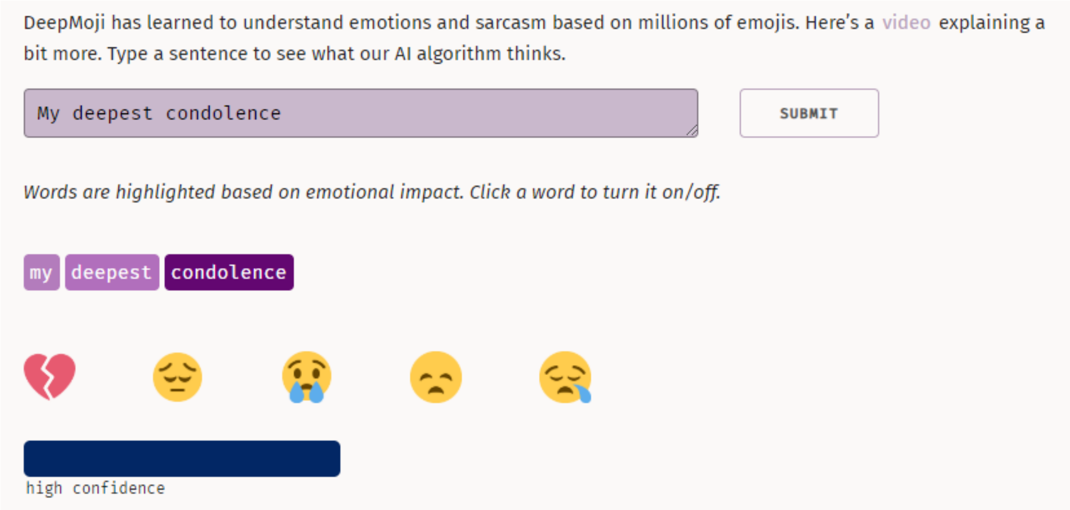


Figure 2.4: Emoji prediction on sad sentences

Machine learning techniques are more important to predict emotion-related emojis. Sentence based prediction cannot decide by a concrete set of rules of computer programming [36]. It is difficult to differentiate the thoughts and expressions of the two people. They may use different techniques to express feelings like happy, sad, serious, or sarcastic [16]. Emotion differs from various external influences such as culture and regions.

MIT deep emoji research focuses at teaching their artificial intelligent agents on human emotions [39]. People may not reflect their real feelings as they were writing. As the MIT deep emoji research describes it is convenient to convey emotion by the word in post and sentences published [37].

2.2.3. Active appearance model

Face detection and facial expression recognition are some of the most challenging concepts [30]. There are several famous techniques for recognizing facial expressions [43]. This paper discusses the approach of the Active Appearance Model (AAM) [6] to human-computer identify real-time facial expressions accurately and quickly with a minimum of delay. In point of fact, AAM [31] has still only been applied to recognize faces and this is an effort to step into another level. Recognition of human to identify the availability in front of the mobile device was a necessity to improve human-computer interaction. Therefore human face detection and facial expression recognition were considered to check the availability of the user.

The face is the main focus of attention in social interaction, which is playing a key role in transmitting identity and emotion. It is quite remarkable that human beings are capable of recognizing thousands of faces learned right through their lifetime and recognize familiar faces at a glance even after years of separation. As the expansion of computer vision, many researchers are very keen on the areas in which they could exploit their findings [8] Real-time facial expression recognition and face detection is regarded as the most common applications of computer vision.

There have been several issues in selecting an appropriate technique that can quickly and precisely recognize facial expressions in order to combine human emotions with 3D emoji icons. Some of the key and most common techniques of facial expression recognition are Facial Action Coding System (FACS) [7] Facial Expression Recognition Techniques Using Constructive Feedforward Neural Networks and K-Means Algorithm Facial Expression Recognition using a Dynamic [30] Model and Motion Energy [11]. But considering the fact that in all of these methods retrieval time becomes much higher due to training through artificial neural networks, there is a need for a new technique with a much quicker retrieval time.

This paves the way to recognize facial expressions through the Active Appearance Model (AAM) [31] which actually is the foundation of face detection. AAMs have been established to be an excellent technique for aligning a pre-defined linear shape model that also has linear appearance disparity, to a previously unseen source image containing that object. It also consists of a statistical model of the shape and grey-level appearance of the object of interest which can generalize to almost any valid example.

An image engages finding model parameters that reduce the difference between the image and the created model example which projected into the image. Nevertheless, this might be a complex problem when there are a large number of parameters. However, it is observed that the displacing of each model parameter from the correct value makes a particular pattern in the residuals. The training process based on recognizing these patterns.

Optical tracking and recognition of faces can be classified into two-dimensional and three-dimensional methods [32]. Two-dimensional methods recognize the characteristics of the face from two-dimensional decompositions and transformations of the image.

Three-dimensional methods consider on coordinates of points on the face such as the shape of the cheeks [9]. Facial Expression Techniques are new advancements that took place after the arising of various Face detection techniques. There are advancements to machine interactions for humans using technologies to enhance human-machine interactions. The greater deal of attention focuses on real-time facial expression recognition rather than face detection since face detection has not been considered as a daunting task anymore.

Face detection and facial emotion recognition are widely used with several techniques. There have been several issues in selecting an appropriate technique that can quickly and precisely recognize facial expressions to select the appropriate emotion-related 3D emoji.

Following methods were focused to detect facial expressions using face detection;
The real-time face detection module
Real-time facial expression recognition (FER)
Active appearance model for expression recognition

Real-time face detection module uses the main reason for choosing the OpenCV library is to get the service of one of the most efficient face detection methods available - Viola Jones Ada- boosted Algorithm [12]. Viola-Jones method uses rectangle filters as masks shown in Figure 2.5: Applying rectangle filters on database images to create an integral image as a new image representation.

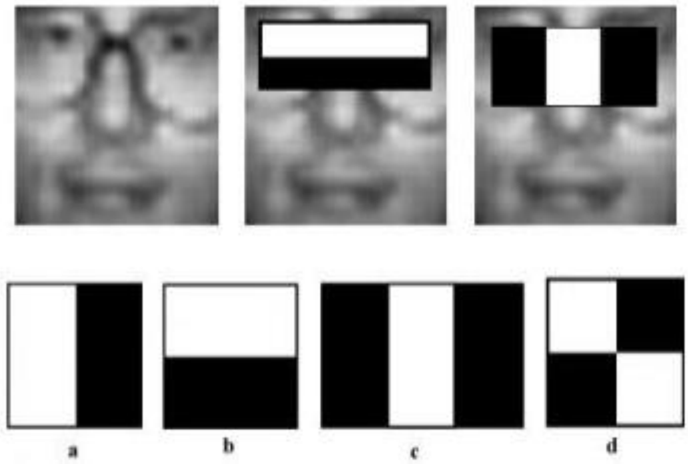


Figure 2.5: Applying rectangle filters on database images (Ngo et al., 2009)

There are many rectangle filters available in the Viola-Jones Ada-boosted algorithm. Each and every sample image in the database has been evaluated with each rectangle filter which ultimately creates an integral image for each rectangle filter-sample image combination.

Afterward, these integral images are sorted according to filter values. However, selecting the best threshold for each filter based on the minimum error is quite a challenge. For this purpose, the sorted integral images are used in order to scan for the optimal threshold. Thereafter selection of the best filter/threshold combination is done and all the sample images are subjected to reweight eventually. Because of the sorted nature of this algorithm, time spends to detect faces becomes pretty much less. Therefore this technique has been regarded as one of the most efficient face detection methods.

Real-time facial expression recognition (FER) is considered to be the most significant and challenging module. The activation of this module begins after the face detection module only when there is a user available in front of the mobile devices that use for online communication.

Figure 2.6: Facial expressions (sleepy, happy, and exhausted)

An active appearance model for expression recognition is another way of detecting human emotions based on other research papers. Active Appearance Model (AMM) one of the most widely used techniques for recognizing faces and tracking [31]. This model locates face feature points in each sample image and then determines expressions according to the variations of those points.

In fact, the Active Appearance Model is a commanding generative class of methods for cases such as modeling and registering deformable visual objects. In recent years this technique has been really admired by experts due to its exceptional performance. The major highlighted feature of AAM is its compact representation of appearance, including shape and texture in addition to its quick fitting to unseen images [6].

The process of recognizing facial expression through AAM is quite a challenge. The input image or web camera video stream is subjected to be tracked by AAM tracker [45], which is computed using a large number of images (Face Database), in order to extract useful facial features. The classification of expressions has been performed using Support Vector Machine (SVM) [10].

Support Vector Machines (SVMs) have been established to be extremely useful in a number of pattern recognition tasks including face and facial expression recognition. This type of classifier endeavors to find the hyperplane that maximizes the margin between positive and negative observations for a particular class.

A linear SVM classification decision is made for an unlabeled test observation o* by, where w is the vector normal to the separating hyperplane and b is the bias. Both w and b are estimated so that they reduce the structural risk of a train-set.

Where

o^* is the unlabeled test observation

w is the vector normal to the separating hyperplane

b is the bias.

Characteristically, w is not defined explicitly, but through a linear sum of support vectors. As a result, SVMs present additional appeal as they tolerate the employment of non-linear combination functions through the use of kernel functions such as the Radial Basis Function (RBF) [10], polynomial, sigmoid kernels. A Radial Basis Function kernel is used in our experiments throughout this paper due to its good performance, and the ability to carry out well in many pattern recognition tasks [10]. Since SVMs are essentially binary classifiers, special steps must be taken to expand them to the multi-class scenario required for facial expression recognition.

SVM obtains the services of one of the SVM methods called "one-against-one" approach (Hsu, Chang & Lin, 2005) in which $K (K - 1)/2$ classifiers are constructed, where K is the number of Action Unit classes, and each one trains data from two diverse classes. In classification, a voting strategy is used, where each binary classification is regarded to be a single vote.

A classification decision is achieved by choosing the class with the maximum number of votes. The process of Facial Expression Recognition (FER) through the Active Appearance Model (AAM) was accurate. Most of the research papers used four poses of the human face into consideration such as exhausted, sleepy, happy and neutral. However, the accuracy of the system varies according to the available lighting conditions. Therefore sometimes it takes a few seconds delay when recognizing expressions.

Facial expression analysis includes both measurements of facial motion and recognition of expression. The general approach to automatic facial expression analysis (AFEA) [13] consists of identifying facial action units or prototypic emotional expressions. Incorporation of face detection with facial expression recognition guarantees a better and meaningful solution for detecting appropriate 3D emoji based on facial emotion category.

Facial expression recognition is one of the difficult and complex task computer sciences is facing. However, there are various methodologies available for solving the task of expression identification such as Facial Action Coding System (FACS), Facial Expression Recognition Techniques Using Constructive Feedforward Neural Networks and K-Means Algorithm and Facial Expression Recognition using a Dynamic Model and Motion Energy.

Although these techniques are able to identify facial expressions, they have their own pros and cons. In fact, for a real-time expression recognition system, less retrieval time is a must. To fulfill this aspect the method of point correlation, Active Appearance Model (AMM) is used to retrieve real-time facial expressions in a short time frame. It ensures results with a quick response time through an acceptable precision.

Both face detection and FER modules have been completely developed with approximate higher accuracy. Actually, it is really hard to achieve 100% accuracy although more than 5000 sample images were trained using q-training and AAM fitting for more than two days. Nevertheless, the face detection module can further be extended to a motion detection module in order to cope up with more advanced practical usages. Accuracy levels can be varied according to the angle and the various lighting conditions.

As a summary of the related others works Table 2.1: Comparison over technologies describes the comparison of each module of the literature survey. The table describes several types of modules and during this literature review communication gap was identified based on facial emotions. The realistic emotions of the user would not be considered majorly on communication. Therefore, based on the literature review, the caption of user facial emotions which changes based on real-time emotion of a three-dimensional emoji on communication was selected.

Table 2.1: Comparison of technologies

Module	Test description	Algorithms and techniques	Issues with the module
Real-Time Face Detection	Checking the state of the monitor whether a user is available or not in front of the mobile device	Viola-Jones Ada-boosted algorithm	Accuracy of face detection decrease based on lighting effects.
Facial Expression Recognition (FER)	Checking the emotions of the user and predict the emotion-related emojis accordingly	The activation of this module begins after the face detection module only when there is a user available in front of the mobile device	Reliability of expression type is low if analyzed using blurred images
Active appearance model	Checking the input image or webcam video stream is subjected to be tracked by AAM tracker Representation of appearance, shape, and texture	Use Face featuring points.	Distortion occurs based on lighting standards and reduced the accuracy level
Google cloud vision API	Extract objects or text in an image	Face detection and emotion recognition	Provides a level for emotion type as integer value rather decimal values.
Deep Emoji	Sentence analysis for emoji prediction	Machine learning related to artificial intelligence	Emoji prediction only relies on text sentences rather than real-time facial expressions

As illustrated in Table 2.1: Comparison over technologies and techniques for emotional emojis can be compared. The comparison table describes Real-Time Face Detection, Facial Expression Recognition (FER), Active appearance model, Google cloud vision API and Deep Emoji on MIT modules. The uses of algorithms and techniques are also summarized and illustrated with respect to test related descriptions and results in each module.

2.3. Summary

This chapter describes the technologies adapted to the research with the reasons for selecting technologies for the system. Mainly the technologies like image processing, facial expression recognition, Google cloud vision API which have been used for the research. The chapter describes similar approaches to the research papers and technologies of usage. The comparison of technologies has been described throughout this chapter. The next chapter describes the approach and design of the suggested research study.

Chapter 3

Design

3.1. Introduction

The previous chapter describes the technologies that have been adapted to solve the research problem. This chapter mainly provides high-level design and architecture about the approach of the research. The detailed description of the approach is also described indicating the users, input, output, features and the methodology of the process.

3.1.1. Methodology

The main approach of the research is to identify the required requirements for proposed research study and functional details to carry out the research. The system architecture was designed by analyzing the information gathered through the literature survey. There were some limitations that were identified in the current approaches. The ideas and information gathered through concept papers were incorporated in designing the system architecture.

To gather requirements and to identify what needed to be incorporated in the research architecture, several existing tools and techniques were analyzed on literature review section.

The system architecture was developed based on these design concerns and identified requirements. The process of communicating each module is considered during the research and focused on evaluation process to evaluate the research.

The approach was identified based on the modules and their capabilities on the research design. The qualitative data and quantitative data were used to analyze and decide the approach of the methodology.

Collecting data, analyzing and integrating information were involved with the use of qualitative and quantitative research information. Quantitative data includes close-ended information such as attitudes, behaviors of humans. The emotional changes of humans are considered as behavior variations for the input of the designed architecture.

Qualitative data is known as open-ended information, where information can be gathered through interviews of people and focus groups of the target groups. Their observations and responses are considered to evaluate the proposed research design.

3.1.1. Design concerns

There are some aspects to be considered to construct the design of the research architecture. The overall system architecture drives through the design components of the system.

One of the concerns is using two-dimensional emojis over three-dimensional emojis. Two-dimensional emojis are widely used in communication systems on social media. Emoji exchanges are more likely to use during communication. As people do not like to share their own faces with expressions, instead of sharing their own faces they share two-dimensional emojis over communication channels.

As per others work and related work on critical literature review, three-dimensional visualization is more effective rather than two-dimensional visuals. Real-time emotion exchanging is also widely used in communication with two-dimensional emojis. Animated three-dimensional emojis are effective and give a realistic feeling during communication with one person or a group of people.

The next concern is the facial emotions of the users. There are many facial emotion types available. It is also a challenging task to detect each and every type of facial emotions of the human face. The real-time facial emotion capturing is another model of the research system architecture.

Google vision API was used for effective and accurate facial emotion detection to eliminate false and wrong facial emotion detection. A video stream was captured while continuing the communication, in order to detect the real-time facial emotions of the users.

Facial features were considered to capture the type of facial emotions using facial expressions. Eyes, eyebrows, mouth, chin and other facial muscle movements are important to decide the human facial emotion type. There is a numerous number of facial emotions are available for human faces. Only a few human facial emotion types were considered for facial detection from human facial expressions.

Emojis are classified based on the type of human facial emotions. The real-time facial emotion capturing was done with the use of the live video stream of the human during communications [40]. Some human faces are detected with emotions by human emotion prediction.

Three-dimensional emojis are classified based on the emotions as well as the skin type of the human. There are researches that completed to define various types of emotions on two-dimensional emojis with a different type of skin variations as it a more variable during communication.

The skin color variation of different types of the region is also considered when animating three-dimensional emojis during communication channels using t-distributed stochastic neighbor embedding (t-SNE) [52]. High-dimensional space clusters, semantically similar emoji together automatically in this space. As an example, faces arrange with the happy in one region, the angry in another. These technologies assist more realistic communication ability for people while exchanging their real-time facial emotions.

3.2. Top level architecture design

The intention is to minimize the emotional communication gap between the message sender and the receiver. Text message communication system can be identified as two subcategories with respect to the number of users as shown in Figure 3.1: Proposed facial emotional exchange. The single user message communication system consists of one text message sender and one text message receiver. The multi-user message communication system consists of a group of the user where each person can receive the same text message.

To achieve this goal exploratory type of research was carried out. Many research documents, tools were explored and analyzed before deriving the proposed system architecture. The research methodology involves both quantitative and qualitative methods while focusing more on qualitative methods. The system architecture is designed with the main aim of improving the quality of extracted conceptual relationships.

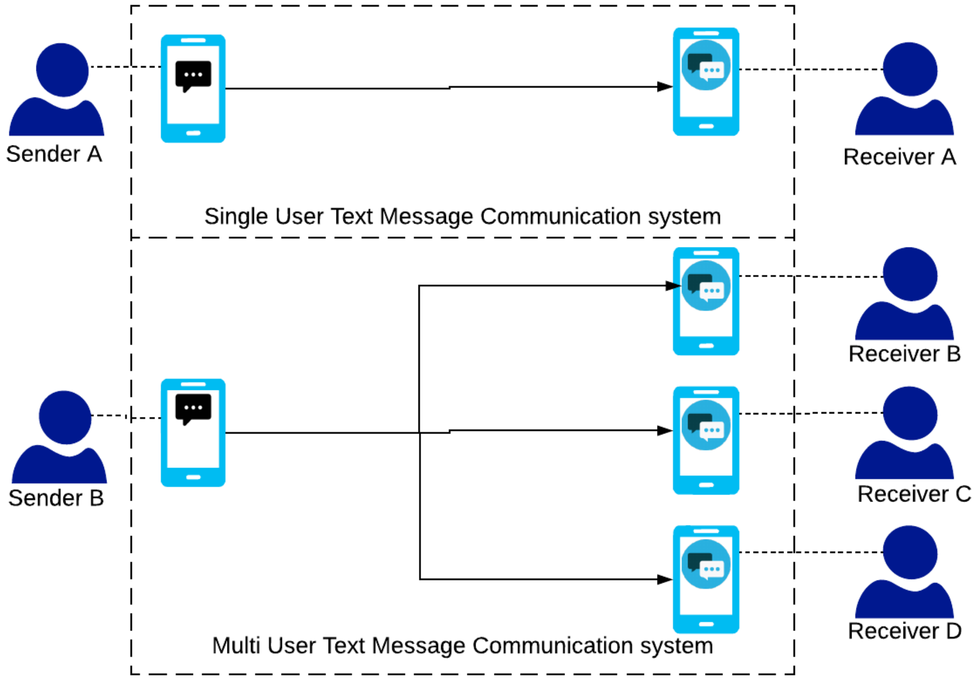


Figure 3.1: Proposed facial emotional exchange

The emotions of the 3D emoji change over time-based on the real-time emotion variation of the user. Therefore, the 3D emoji emotion classification is an essential section as the behavior of the emoji should be adopted based on the user's interaction with the text message communication system. As shown in Figure 3.2: 3D emoji classification, the text message sender is the single person and the face detection should happen to check the state of a real human is available in front of the mobile device. Feature extraction, facial feature classification, and expression recognition methodologies follow after the face detection method is successfully completed. The laughing face of the user is expressed to another party in the communication system via 3D emoji with a laughing animation to denote the real-time behavior of the message sender.

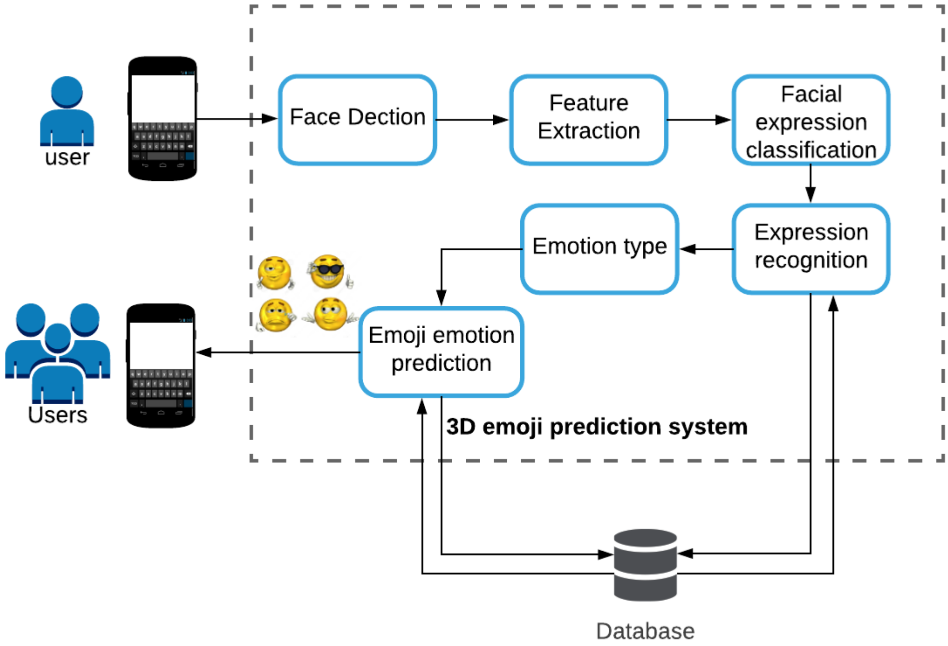


Figure 3.2: 3D emoji classification

2D emojis are widely used in text message communication to denote the emotions that cannot be denoted through text message. As per the researches, 3D is an essential and effective method to perceive emotions and behavior. The most efficiently 3D objects are capable to perceive real-time realistic behaviors and interactive effect for communication as text messages cannot express all the feelings thoughts and emotions to the other party of the text message passing system.

The message communication system needs to recognize the availability of the user and decide the rest of the activities in the system. Figure 3.3 shows the face detection from the live video stream from the message sender through a mobile device. Eye recognition is a static way to define a human face without a fake object. The face detection method and the expression recognition followed by completing the eye and face tracking phase [32].

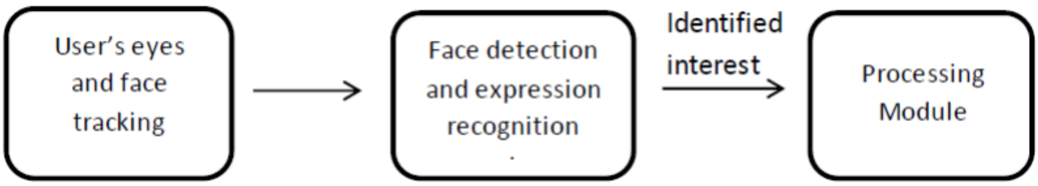


Figure 3.3: Face detection and processing module interaction

Emotion recognition of the human face can be conducted via API's given by Google cloud vision. Facial expression and emotion classification is capable of this tool and feasible to easily integrate with other tools. The 3D emoji models use the extracted facial feature changes and emotion variations of the text message sender and adopt the emotional variations to 3D emoji emotion varying techniques.

3.3. Approach for the proposed solution

The research paper is about Facial Emotion Recognition Using Facial Expression. The method of Facial feature classification is to identify facial emotions. Identify classification algorithms for facial emotion recognition [1]. The method of facial emotion recognition (FER) [18] can be classified as input images using video clips, face detection, feature extraction and facial expression classification [17]. Emoji is used to express emotional feelings to other parties. Therefore emoji classification will be conducted to investigate emotional recognition using a facial expression by emoji [4].

Real-time interactive communication between gives users a realistic feeling. As adhere to the methodologies graphical changing ability of 3D emoji should be adopted to complete research on real-time graphical changes on 3D emoji. The below Figure 3.4: Sequence diagram of the system shows a sequence diagram of the proposed solution and the interaction between the components is also shown to describe the workflow.

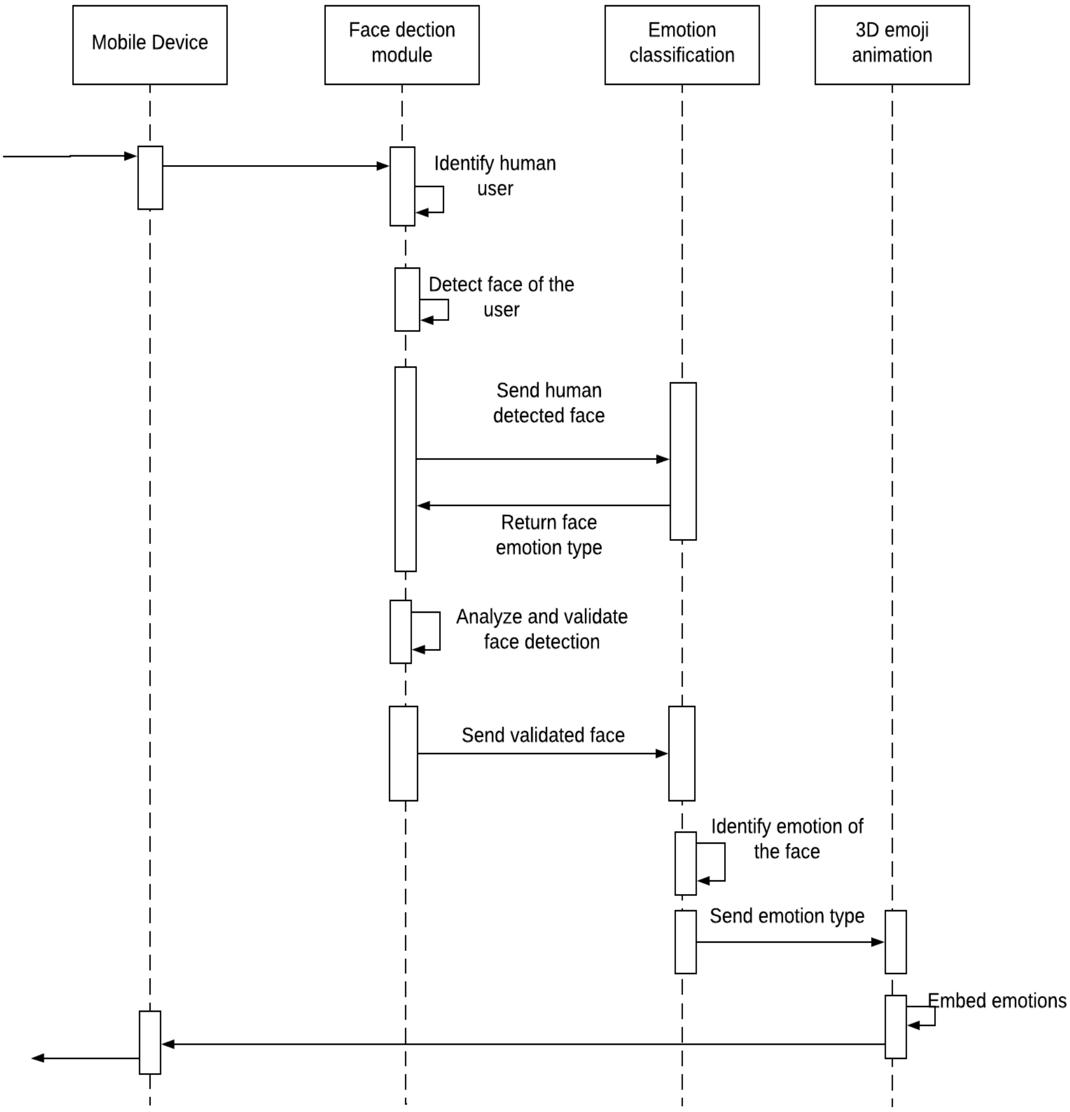


Figure 3.4: Sequence diagram of the system

3.3.1. Users

Communication needs two different parties for successful communication. There is a one to one mapping of users or one to many mapping of users. Communication can occur as a single person to another single person or single person to a group of people in a group communication chat application. The target group of the research is more than five years old as the infants cannot actively participate in the proposed research study. The sample size of forty outputs were considered to evaluate the proposed research.

3.3.2. Input

The main input type is the facial emotion of the user who uses mobile devices for chat system based communication. The face detection is the main objective that binds the facial emotions with the face and classifies based on the facial emotions of the user. A video stream of the user is used as the input of the research study. The input video stream, face detection, and facial emotion recognition should more accurate in order to obtain the best results of emoji types.

3.3.3. Output

The output can be visible to a single user or multiple users of the message receiving party on a chat based communication system. The output is based on the facial emotion type of message sending party. The real-time facial emotion of the message sending user can be felt to the message receiving party based on the emoji response. The user facial emotion can be adapted to 3D emoji icons and send real-time feelings and emotional messages to the message receiving party in the communication channel.

3.3.4. Features of the system

Facial emotion recognition section is a challenging method as the facial emotions are very tenuous features in the human face. Emotion can differ from one person to another person depending on their feature changes. Google vision API can be used to analyze images [41] with human faces and analyze emotion types strength based on the feature positions and a learning system.

The face of the user is captured with the use of live video stream and face detection, feature extraction and facial expression classification, facial expression recognition methodologies are conducted using Google cloud vision API as most of the facial feature classifications are available.

Exchanging feelings and emotions of each communication party are the most important feature to solve the communication gap via text message communication applications. The emotional changes can be done to 3D emoji icons where text message exchange parties can send to each other and express emotional feelings with each other. Emotional changes of the emojis can be adopted based on the user's emotions and a more realistic feeling is given to text message exchange parties. The facial feature of 3D emojis can be changed based on the real-time facial expressions of the user.

3.4. Design assumptions

Emotions of the user has to be detected to decide the emotion category of the emoji. This classification structure suggests that there is an assumption of accuracy of the video stream of human face. There can be many noises and distortions are being added for the input data which reduce an accurate result of the experiment. Therefore, it is assumed that the noises and distortions have been discarded to get a real solution with input for the solution.

Detection of human face from a live video stream is a challenge for the solutions. Therefore the accuracy level can be detected with the real-time facial emotion recognition. The assumption is that frequency level of the image on a video was reduced to a few frames detection on the video to identify the human face. Therefore low numbers of image frames were considered to calculate and detect the facial emotion type category.

Predicting facial emotion category was decided based on the facial feature extraction. Lip positions, eye brows location, eye pupil position and many more features are used to detect the facial emotion types using facial expression. All the useful facial feature categories are considered when identifying the facial emotion type.

3.5. Summary

This methodology chapter describes the approach to conducting the research as part of the methodology adopted. The systematic approach has been discussed in this chapter. The next chapter describes the implementation of the proposed research.

Chapter 4

Implementation

4.1. Introduction

To denote the research design and approach a prototype system was developed for the purpose of completion of the research. This chapter describes the implementation issues, challenges encountered and decisions made to achieve objectives.

4.2. Face detection

There is a way to get started with human face recognition using Python [47] as well as the open source library called OpenCV. OpenCV is a library for computer vision. This is originally written in C/C++ and feasible of bindings for Python [46].

There are inbuilt capabilities in OpenCV to recognize faces in an image for face detection. Therefore this process is conducted using openCV, with the use of machine learning algorithms to search for faces within a picture [48]. Because faces are so complicated, there are many ways to detect if it finds a face or not.

Therefore there pre-implemented thousands of small patterns and features to refuse to check the matching items. The algorithms decompose the task of identifying the face into thousands of smaller, bite-sized tasks, where it is easy to solve. So that classifiers are important to define tasks in the algorithm. The library uses multiple cycles of stages to detect the faces in a picture.

During the critical analysis of literature review conducted, Google vision API is one of the best applications for image analysis. This application programming interface is easy to use and accurate in detection faces in an image. Therefore, the images extracted from the live video stream is able to classify in to face detection with several types of content.

4.3. Facial emotion extraction

The facial expression recognition approaches have generally attempted to recognize the expression of six basic emotions named as happiness, surprise, sadness, disgust, fear, and anger. Emotion detection using facial expression is a sort of a challenging task considering the accuracy of human facial emotion variations.

There are application programming interfaces to extract emotions from human facial expressions. Some of the application programming interfaces are Google vision API and Kairos API and there are more applications available as well. Considering the applicability and usability of the application Google vision API was used to detect the facial emotions of the human face.

Google vision API can be used to detect the face and its emotion at the same time. The content of the image is extracted and the facial emotions are predicted based on the facial feature positions. Eyes, eyebrows, chin, and other facial muscle location and positioning is analyzed to extract the facial emotion type. As shown in Figure 4.1: Happy facial expression detection, vision API is capable of detecting the joyful face based on the facial features in the face are positioned. The strength of the prediction of facial emotions is shown to understand the gravity of facial emotion using facial feature extraction.

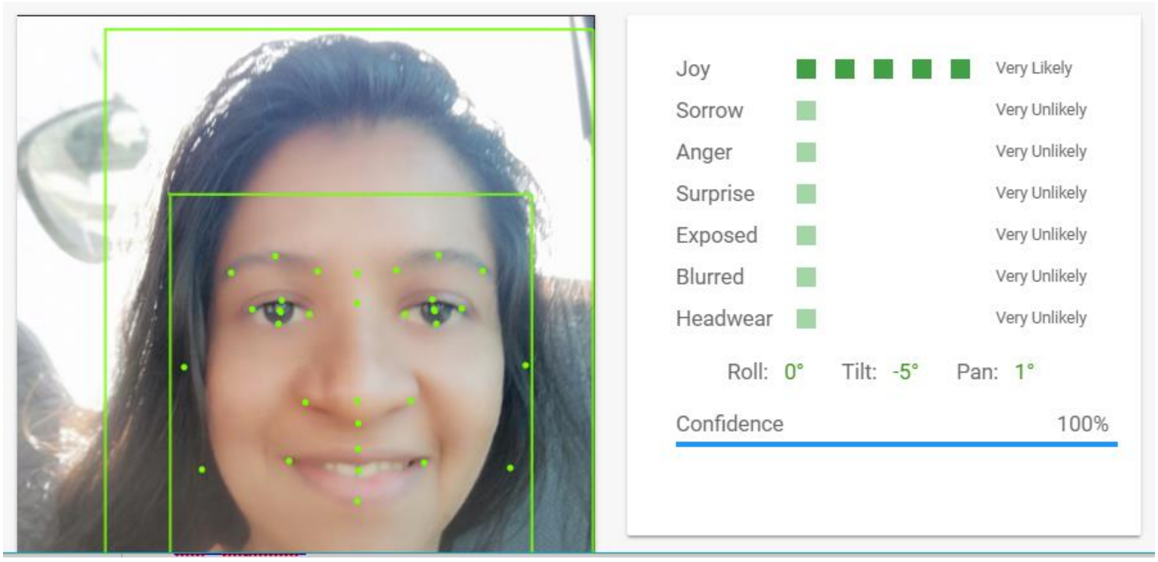


Figure 4.1: Happy facial expression detection

The output data is shown in a different type of formats where more readable for human users. As shown in Figure 4.2: Happy facial expression detection JSON output, there is JSON output where the application can easily deserialize the JSON output data and embed information into three-dimensional emojis.

Figure 4.2: Happy facial expression detection JSON output

Both Figure 4.3: Surprised facial expression detection and Figure 4.4: Image request and response shows another type of facial emotions that were identified using the Google vision API. Based on the facial features the strength of the detected emotion type is low. This is a sort of a challenging task, where the accuracy of the facial emotion detection occurs.

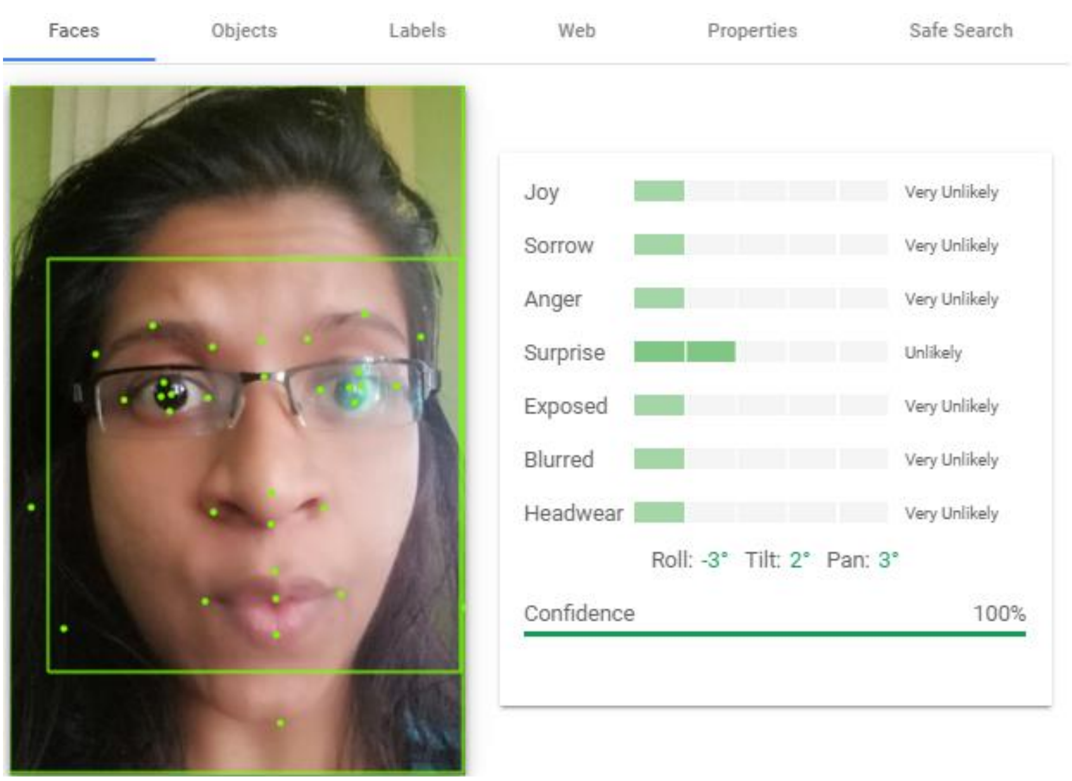


Figure 4.3: Surprised facial expression detection

Figure 4.3 expresses a surprising emotion of a human face. The position of the mouth eyes and eyebrows are considered to decide the surprised face and the application programming interface is capable of deciding the strength of the emotion extraction.

Figure 4.4: Image request and response

If a sad emotion of a human face is added to the system, the same approach of the position of the mouth eyes and eyebrows are considered to decide the emotion type of the face. Figure 4.4: Image request and response, shows the request data type and the dedicated response about the emotion type in the image.

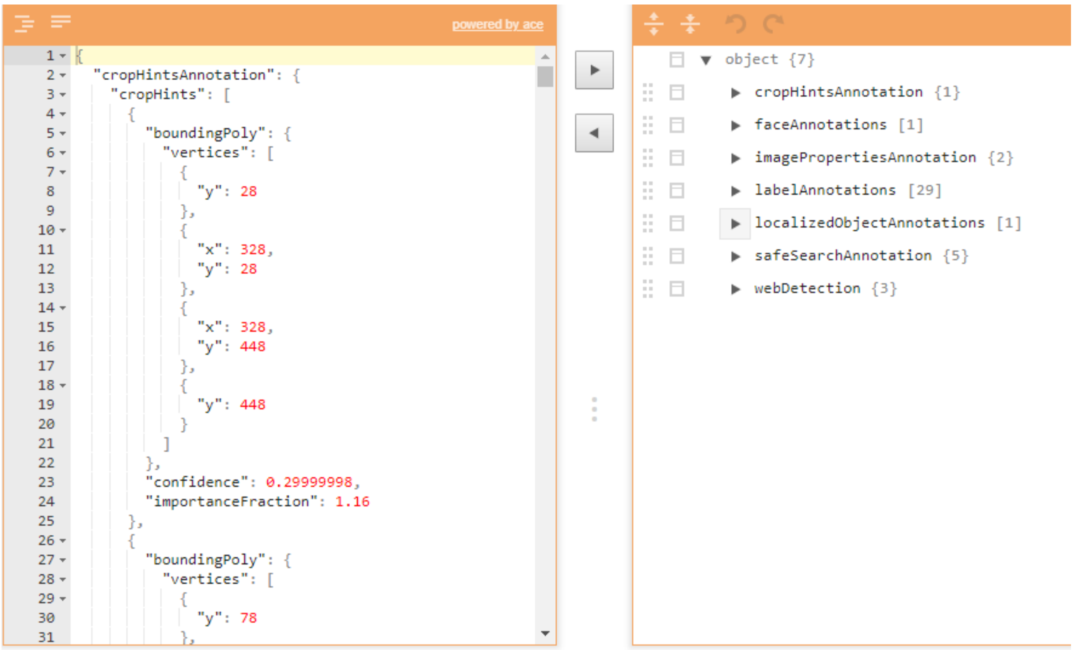


Figure 4.5: Format of the response object

The application programming interface is used to decide the strength of the emotion extraction. The detected facial emotion is used to embed into the three-dimensional emoji for effective and realistic communication in social media and other communication channels [27]. As shown in Figure 4.5: Format of the response object, there are multiple data formats available in response taken from the request data in the image. faceAnnotations, imagePropertiesAnnotation, and labelAnnotations are some of the sections that hold facial emotion details in the response. Facial landmarks like eyelashes, eyebrows, forehead, nose, cheek, lip movement [21] and chin are some of the main facial features that helped to detect the type of facial emotion using facial expressions.

The prototype of the research study was built to analyze the proposed architecture. The results were taken from a group of people for analysis purposes. The age level and gender were considered to obtain the output from the prototype of the research study. The scoring mechanism is used for rating each scenario. The score value from 1 to 5 were used rate and the wording of each level was used to analyze the results obtained.

Scoring values are as follows:
1: Very dissatisfied
2: Dissatisfied
3: Neutral
4: Satisfied
5: Very satisfied

Table 4.1: Rating of sample users

Age group	Gender	True positive	True negative	False positive	False negative
20-25	Male	2	NA	NA	NA
20-25	Female	NA	NA	4	NA
20-25	Male	4	NA	NA	NA
20-25	Female	4	NA	NA	NA
20-25	Male	NA	NA	2	NA
20-25	Female	NA	NA	NA	NA
20-25	Male	NA	NA	NA	4
20-25	Female	5	NA	NA	NA
20-25	Male	4	NA	NA	NA
20-25	Female	2	NA	NA	NA
20-25	Male	3	NA	NA	NA
20-25	Female	NA	NA	2	NA
20-25	Male	NA	4	NA	NA
20-25	Female	NA	NA	NA	3
20-25	Male	NA	4	NA	NA
20-25	Female	5	NA	NA	NA
20-25	Male	4	NA	NA	NA
20-25	Female	2	NA	NA	NA
20-25	Male	4	NA	NA	NA
20-25	Female	2	NA	NA	NA
25-30	Male	1	NA	NA	NA
25-30	Female	2	NA	NA	NA
25-30	Male	5	NA	NA	NA
25-30	Female	NA	NA	NA	2
25-30	Male	5	NA	NA	NA
25-30	Female	2	NA	NA	NA
25-30	Male	NA	NA	4	NA
25-30	Female	2	NA	NA	NA
25-30	Male	2	NA	NA	NA
25-30	Female	NA	2	NA	NA
25-30	Male	2	NA	NA	NA
25-30	Female	2	NA	NA	NA
25-30	Male	4	NA	NA	NA
25-30	Female	NA	NA	4	NA
25-30	Male	NA	NA	NA	4
25-30	Female	NA	NA	2	NA
25-30	Male	4	NA	NA	NA
25-30	Female	NA	4	NA	NA
25-30	Male	2	NA	NA	NA
25-30	Female	5	NA	NA	NA

As shown in Table 4.1: Rating of sample users, people for the prototype were selected based on different age groups. Gender also considered when taking ratings from the proposed research study. Rating 1 describes the scoring level as very dissatisfied for the testing scenario while rating 5 describes the scoring level as very satisfied for the scenario. This satisfaction levels of users are considered as results of the proposed prototype of the research study.

4.4. Summary

This implementation chapter describes the implementation plan to guide the design of the solution. The implementation related systematic approach has been discussed in this chapter. The next chapter describes evaluation of the proposed research.

Chapter 5

Evaluation and results

5.1. Introduction

The previous chapter describes the high-level design and the approach of the research project. This chapter mainly provides details of the evaluation criteria of the research. The evaluation methodology is essential for the qualitative analysis of the research.

5.2. Evaluation plan of research study

Summarizing real-time interaction success ratio from users by graphical representation of 3D emoji is essential to evaluate and decide the success of the system. Analyze face detection from the real-time video stream should be accurate enough to identify the emotions of the single user. The research scope limits to a single user to identify relevant facial emotion of the user and emotional changes are embedded to 3D emojis with animated graphical variations. Identification of relevant 3D emoji for the specific facial emotion on the image of video stream should be more accurate.

Evaluation process conducted via a selected set of parameters gained by the research methodologies. Inputs of the system can be considered as one section of the evaluation process of the research system. In the mentioned input values are the video stream for mobile devices with a text message communication system. As the evaluation process is to monitor the accuracy and the efficiency of the system it is important to identify the particular most effective frames among the set of frames in the input video stream.

Another type of evaluation process depends on the dependent variable that influenced by the independent variable. The dependent variable is the emotion type categories which vary according to the independent variable of the detected face. A ratio is useful to count the success rate and failure rate due to the variable types and combination of two types of variables.

Evaluation can be conducted based on the 3D emoji categorization related to the emotion of the user who uses the text messaging passing system. The accuracy level and effectiveness level of the emotional change on 3D emoji should be calculated to evaluate the research system to maintain a better relationship with the input of the system and expected output.

5.3. Evaluation from precision and recall

Precision-recall graphs and F-measure have been used in order to evaluate the results of the proposed module. Precision called as positive predictive value is the fraction of retrieved instances that are relevant, while recall which is also known as sensitivity is the fraction of relevant instances that are retrieved [29].

For classification tasks, the terms true positives, true negatives, false positives, and false negatives have been used to compare the results of the classifier under test with trusted external judgment [29]. The terms positive and negative refer to the classifier's prediction which is known as the expectation, and the terms true and false refer to whether that prediction corresponds to the external judgment which sometimes known as the observation [20].

Table 5.1: Classification tasks

Symbol	**Description**
tp	True Positive (TP)
tn	True Negative (TN)
fp	False Positive (FP)
fn	False Negative (FN)

True Positive (TP) – Equivalent Hit
True Negative (TN) – Equivalent with correct rejection
False Positive (FP) – Equivalent with false alarm
False Negative (FN) – Equivalent with miss
Precision and recall are then defined as [20].

$$\text{Precision} = \frac{tp}{tp + fp}$$
$$\text{Recall} = \frac{tp}{tp + fn} \tag{5.1}$$

Usually, precision and recall scores are not discussed in isolation. Instead, values for one measure are compared with other measures. Precision and recall values are calculated with the use of classification tasks (*equation 5.1*). Then the results were evaluated through precision-recall-graph.

Precision and recall can be combined to produce a single metric known as F-measure, which is the weighted harmonic mean of precision and recall where an F1 score reaches its best value at 1 and worst score at 0. In the statistical analysis of binary classification, the F1 score (also F-score or F-measure) is a measure of a test's accuracy (*equation 5.2*).

$$F = 2 \cdot \frac{\text{precision} \cdot \text{recall}}{\text{precision} + \text{recall}} \tag{5.2}$$

In the following sections, the precision-recall graph will be plotted and F score will be calculated for the module. Then the accuracy of each module was further discussed, considering the F score value.

5.4. Evaluation procedure

The precision of the face detection system can be analyzed as recognizing user interest as looking into the camera or looking away from the camera. The precision of recognizing user interest can be analyzed as below.

True Positive (TP): User looks into the camera and system identify as it is and recognize emotion type with relevant three-dimensional emoji.

True Negative (TN): User looks away from the camera and it is being recognized as looking away and three-dimensional emoji prediction does not continue.

False Positive (FP): User looks away from the camera and the system identifies it as looking into the camera where irrelevant three-dimensional emojis have predicted.

False Negative (FN): User looks into the camera but the system does not recognize it as looking into the camera and any predicted emojis does not predict.

The above criteria are the measurements to decide the specific functions and their accuracy. The true positive is the time of user looks into the camera and system identify as it is and recognize emotion type with relevant three-dimensional emoji. Several rounds of specific levels are being considered to calculate the result. True negative measurement is obtained using at the time the user looks away from the camera and it is being recognized as looking away and three-dimensional emoji prediction does not continue.

In order to calculate the precision and recall false positive measurement is also considered as the user looks away from the camera and system identifies it as looking into the camera where irrelevant three-dimensional emojis have predicted. Values are required for this section by trying out several times of the system. False negative opportunity can be taken when the user looks into the camera, but the system does not recognize as looking into the camera and any predicted emojis does not predict.

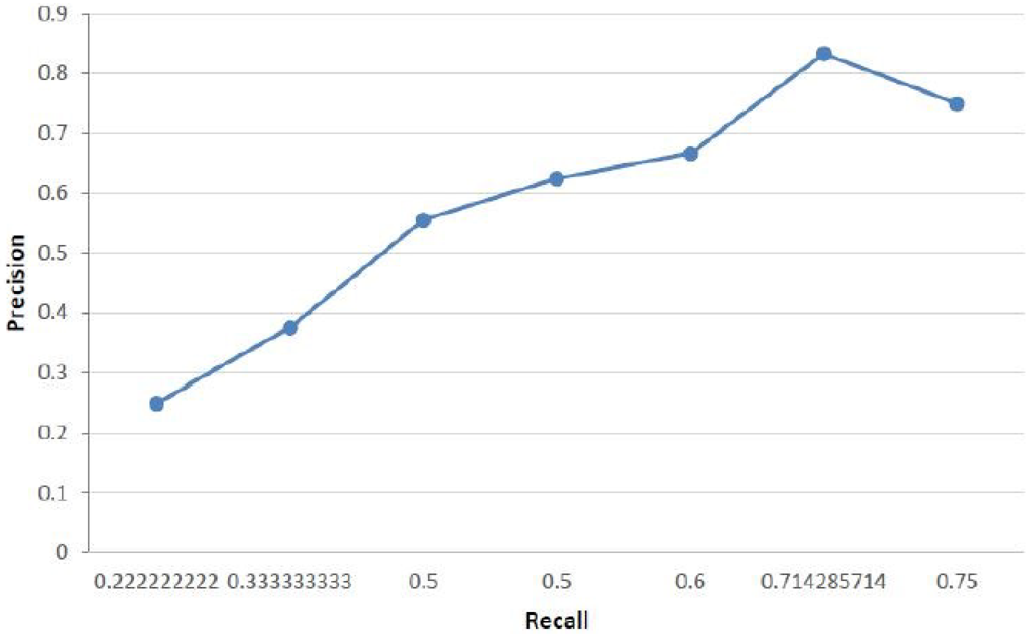


Figure 5.1: Precision vs. recall graph for the emotion recognition

Precision and recall values differ for several system configurations and different conditions that setup when evaluating the system for the proposed solution. For the calculations of F1 Score, average values of Recall and Precision have been calculated. Therefore the percentage of the accuracy level is calculated and based on the precision and recall values

For the calculations of F1 Score (*equation 5.2*), average values of Recall and Precision has been
calculated.
Average Recall = 0.58832
Average Precision = 0.63741

F1 = 0.61188
It can be seen from the above result that the F1 Score for the scenario; recognizing user looking into the camera or looking away from the camera, is 0.61188.So that as a percentage this scenario has 61% accuracy. The F1 score gives the average percentage of the level of succession with the proposed solution.

5.4. Summary

This chapter provides an evaluation of the proposed research. The chapter describes the importance of the evaluation process and results of the data analysis. Average recall value was calculated as 0.58832 and average precision value was calculated as 0.63741. Based on the evaluation, F1 score was calculated and 61% percent of accuracy level was obtained based on the results of sample group. The next chapter describes the discussion and conclusion the research study

Chapter 6

Discussion and conclusion

6.1. Introduction

People do not like to share their own real-time videos on social media to let others about their feelings even if they like to share their thoughts while communicating with a group of people. Facebook, Instagram, and Twitter are some of the social media where various types of human emotions are being shared. Real-time facial emotion recognition using facial expressions helps to embed three-dimensional emojis with relevant facial expression type. This thesis investigates human facial emotion recognition using facial expressions and integrating real-time facial emotions with three-dimensional emojis.

People used to share their emotional expressions using two-dimensional emojis rather than sharing their own face on social media. People feel more comfortable sharing two-dimensional emojis without sharing their live facial emotions using a video stream. This thesis examines the realistic feeling of human emotional expression exchange via three-dimensional emojis rather than sharing two-dimensional emojis.

6.2. Conclusion

The research was conducted using constructive research methodology with a practically relevant problem which also has research potential. At the first step understanding of current state of research was conducted by real-time data analysis. This research study has theoretical connections and the research contribution for the solution concept.

The research is based on three-dimensional emoji prediction using facial emotion recognition. There are three basic modules identified to complete the research study with relevant responses as accomplishments. Modules of the research design consists of face detection module, emotion classification module and three-dimensional emoji prediction module. Face detection conducted using face detection module by analyzing a real-time video stream. The face detection was more accurate based on the lighting standards of the environment.

Facial expression extraction for facial emotion recognition is the next step of the research study. Emotion classification module uses the detected human face to extract the human facial expression with facial emotions. Google vision API was used to investigate the emotion type from the detected face of the real-time live video stream. Three-dimensional emoji prediction module consumes extracted facial emotion type to predict associated three-dimensional emoji to the user.

6.3. Future work

This thesis is a piece of clear evidence to determine how effective is to user three-dimensional emojis based on real-time human facial expression variation. It is more realistic to use three-dimensional emojis with human facial emotional changes rather than sharing two-dimensional emojis on social media. The accuracy level of human emotion recognition and the embedding emotional variations into three-dimensional emojis have to be increased. Real-time facial emotional recognition can be conducted with a more accurate manner. This concept can be experienced on almost all social media is a way to improve the live realistic communication while on social media.

Appendix A

About the survey

The thesis is based on using three-dimensional emojis with facial emotion variations, rather than using two-dimensional emojis to express user emotions on social media. The focus age group is for the users who were above 16 years old. The target generation is the people who use social media to express and share their emotional situations on the networks.

Questionnaire for user

- Age (in years)?
- Gender?
  - Male
  - Female
- Are you a software developer or have you been a software developer?
  - Yes
  - No
- Computer hours per week (hours)?

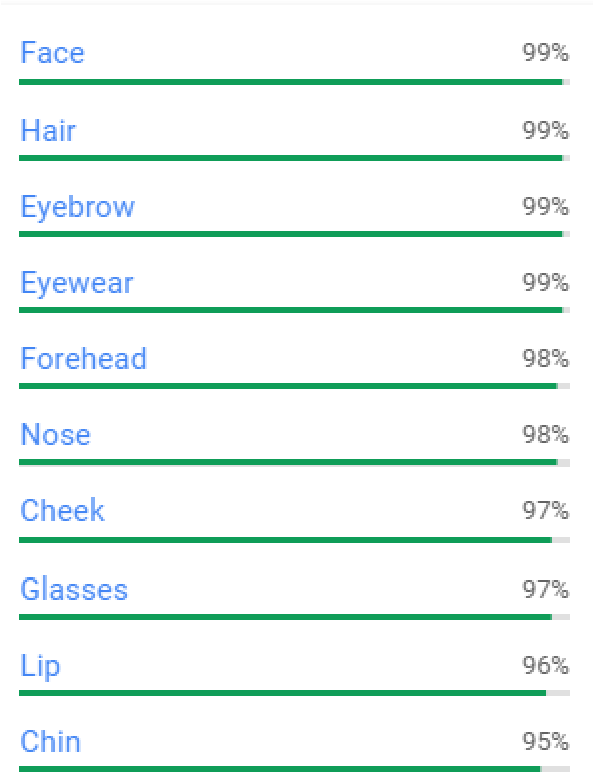
- Do you prefer social media?
  - Yes
  - No
- Do you like to share your feelings on social media? Yes / No
  - Yes
  - No
- Can you express your thoughts and feelings only with words on social media?
  - Yes
  - No
- what do you prefer to share your feelings?
  - Text
  - Emoji
  - Image
  - Other

- Do you like to watch others’ posts based on emotional content?
    - Yes
    - No
- Do you like to share your own image to express your feelings on social media?
    - Yes
    - No
- Do you like to share your video to express your feelings on social media?
    - Yes
    - No
- What type of images do you prefer to get a genuine feeling?
    - Two-dimensional images
    - Three-dimensional images

- What will give you a more realistic feeling?
    - Two-dimensional emoji
    - Three-dimensional emoji

Appendix B

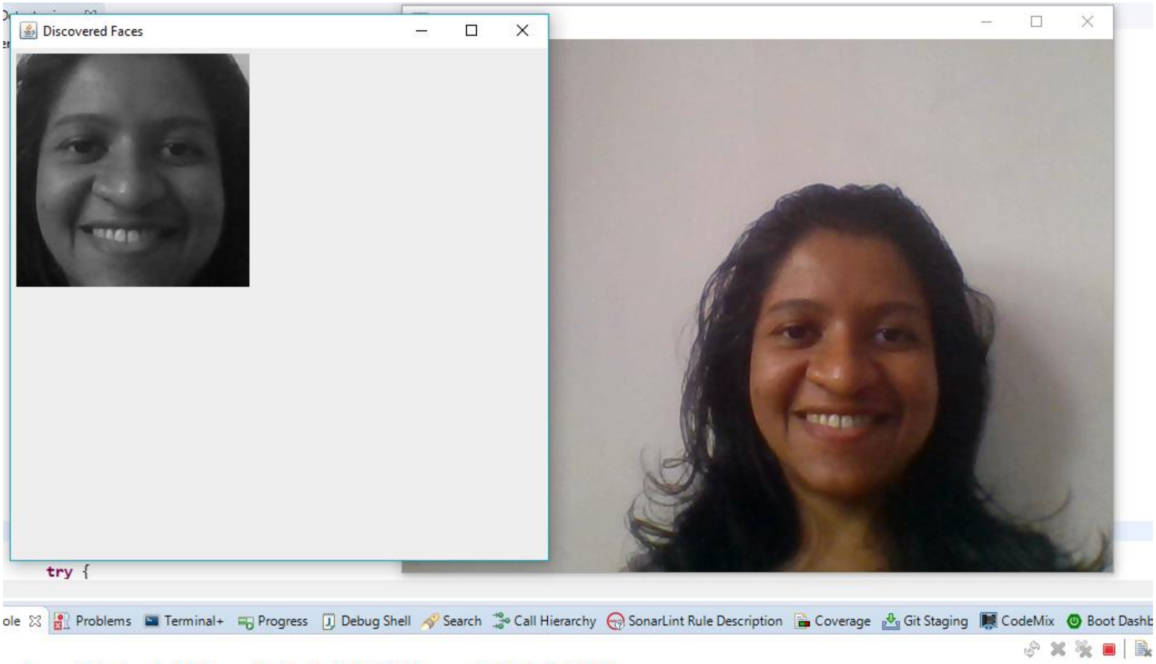
Labels to detect a face with emotions



Request data to manipulate


```
{
 "requests": [
  {
   "features": [],
   "image": {“content": "(data from msc2.PNG)”},
   "imageContext": {
    "cropHintsParams": {}
   }
  }
 ]
}
```


Face detection module



Facial emotion recognition

```
pis.com/172.217.27.42:443] INBOUND PING:
Detection_confidence  3 >  0.5254194
Joy   : Image 3 > VERY_LIKELY
Joy int : Image 3 > 5
Sorrow  : Image 3 > VERY_UNLIKELY
Sorrow  int : Image 3 > 1
Anger  : Image 3 > VERY_UNLIKELY
Anger  int : Image 3 > 1
```

Response from the request


```
{
  "cropHintsAnnotation": {},
  "faceAnnotations": [],
  "imagePropertiesAnnotation": {},
  "labelAnnotations": [],
  "localizedObjectAnnotations": [],
  "safeSearchAnnotation": {
    "adult": "UNLIKELY",
    "medical": "VERY_UNLIKELY",
    "racy": "POSSIBLE",
    "spoof": "VERY_UNLIKELY",
    "violence": "VERY_UNLIKELY"
  },
  "webDetection": {}
}
```

References

[1]. Sail.usc.edu. (2018). Analysis of Emotion Recognition Using Facial Expressions, Speech, and Multimodal Information. [online] Available at: https://sail.usc.edu/publications/files/Busso_2004.p

[2]. Dubey, M. and Singh2, P. (2016). Automatic Emotion Recognition Using Facial Expression: A Review Monika Dubey1, Prof. Lokesh Singh2 1Department of Computer Science & Engineering, Technocrats Institute of Technology, Bhopal

[3]. C. Dhargave, 1., Sonak, 2. and Jagtap, 3. (2015). Survey paper on methods of emotion detection from still images. 1,2M.E. Student, CE Dept, 3Assistant Professor, CE Dept.

[4]. Rajhi, M. (2017). Emotional Recognition Using Facial Expression by Emoji in Real Time. International Journal of Innovative Research in Computer and Communication Engineering.

[5]. Zhou, R., Hentschel, J., and Kumar, N. (2016). Goodbye Text, Hello Emoji: Mobile Communication on WeChat in China. School of International Affairs School of Interactive Computing Georgia Institute of Technology: School of Interactive Computing Georgia Institute of Technology

[6] Akshay Asthana, Jason Saragih, Michael Wagner & Roland Goecke (2009), Evaluating AAM Fitting Methods for Facial Expression Recognition. pp 1-5.

[7]Cootes T, Edwards G, and Taylor C. Active Appearance Models, PAMI, vol. 23, no. 6, pp. 681–685.

[8]Crowley J. L., Coutaz J., & Bérard F. Perceptual user interfaces: things that see. Commun. ACM, 43 (3), 54–64.

[9] Dinges DF, Rider RL, Dorrian J, Mcglinchey EL, Rogers NL, Cizman Z, Oldenstein SK, Vogler C, Venkataraman S, Metaxas DN. Optical computer recognition of facial expressions associated with stress induced by performance demands. Aviat Space Environ Med 2005; 76(6, Suppl.):B172–82

[10] Hsu C., Chang C., & Lin C. A practical guide to support vector classification. Technical report, Department of Computer Science, National Taiwan University

[11]Moghaddam B, Pentland A. Face recognition using view-based and modular Eigenspace, In Automatic Systems for the Identification and Inspection of Humans, volume 2277. SPIE

[12] Viola-Jones & Morphology-based Face Detector, Retrieved from http://www.cc.gatech.edu/~kihwan23/imageCV/Final2005/FinalProject_KH.htm

[13] Ying-Li Tian,1 Takeo Kanade2, and Jeffrey F. Cohn3, Facial Expression Analysis

[14] Google Cloud. (2018). Vision API - Image Content Analysis | Cloud Vision API | Google Cloud. [online] Available at: https://cloud.google.com/vision/.

[15] Google Cloud. (2018). Drag and drop | Cloud Vision API Documentation | Google Cloud. [online] Available at https://cloud.google.com/vision/docs/ drag-and-drop.

[16] MIT Media Lab. (2018). Project Overview ‹ DeepMoji – MIT Media Lab. [online] Available at https://www.media.mit.edu/projects/deepmoji/overview.

[17] Chul Ko, B. (2018). A Brief Review of Facial Emotion Recognition Based on Visual Information. Department of Computer Engineering, Keimyung University, Daegu 42601, Korea.

[18] Shukla, P. and Patil, S. (2018). Survey Paper on Emotion Recognition. International Journal of Engineering and Applied Sciences (IJEAS) ISSN: 2394-3661, Volume-3, Issue-2, February 2016.

[19] Duncan, D., Shine, G. and English, C. (2018). Facial Emotion Recognition in Real Time.

[20] Saito, T. and Rehmsmeier, M. (2015). The Precision-Recall Plot Is More Informative than the ROC Plot When Evaluating Binary Classifiers on Imbalanced Datasets.

[21] G. Zoric, K. Smid and I. Pandzic, 'Automated Gesturing for Virtual Characters: Speech-driven and Text-driven Approaches', JMM

[22] H. Zhao, Emotion-driven interactive storytelling

[23] P. Peter, B. Sylwia, H. Radoslaw, N. Catherine and P. Marc, 'Emotional Interactive Storyteller System', International Conference on Kansei Engineering and Emotion Research, Paris

[24] Stark, L. and Crawford, K. (2015). The Conservatism of Emoji: Work, Affect, and Communication. Social Media + Society

[25] Bazarova, N., Cosley, D., Whitlock, J., Choi, Y. and Sosik, V. (2015). Social Sharing of Emotions on Facebook: Channel Differences, Satisfaction, and Replies. Vancouver, BC, Canada.

[26] Yu, D. and John-Baptiste, S. (2016). Emotion Expression on Social Networking Sites. A Study of Young Persons' Use of Facebook and Twitter in the UK.

[27] Panger, G. (2017). Emotion in Social Media. University of California, Berkeley.

[28] Lin, H., Tov, W. and Qiu, L. (2014). Emotional disclosure on social networking sites: The role of network structure and psychological needs. Computers in Human Behavior, 41, pp.342-350.

[29] Saito, T. and Rehmsmeier, M. (2016). Precrec: fast and accurate precision–recall and ROC curve calculations in R. Bioinformatics, 33(1), pp.145-147.

[30] Neeraj Agrawal, Rob Cosgriff and Ritvik Mudur, Mood Detection: Implementing a facial expression recognition system

[31] Simon Lucey., Ahmed Bilal Ashraf., & Jeffrey F. Cohn Carnegie Mellon University, USA, Investigating Spontaneous Facial Action Recognition through AAM Representations of the Face, pp 2-5

[32] Vacchetti L., Lepetit V., & Fua P. "Stable real time 3D tracking using online and offline information", IEEE Trans. PAMI, vol. 26(6)

[33] Catalin Coman, A., Zara, G., Nechaev, Y., Barlacchi, G. and Moschitti, A.Exploiting Deep Neural Networks for Tweet-based Emoji Prediction. University of Trento, Trento, Italy.

[34] Felbo, B., Mislove, A., Søgaard, A., Rahwan, I. and Lehmann, S. Using millions of emoji occurrences to learn any-domain representations for detecting sentiment, emotion and sarcasm. Media Lab, Massachusetts Institute of Technology.

[35] Karthik, V., Nair, D. and J, A. (2018). Opinion Mining on Emojis using Deep Learning Techniques. Procedia Computer Science, 132, pp.167-173.

[36] Ribeiro, A. and da Silva, N. (2018). Emoji Prediction in Tweets. Institute of Informatics Federal University of Goias.

[37] Rahwan, I. (2019). Using emoji to teach software sarcasm and slang – MIT Media Lab.

[38] Knight, W. (2019). An algorithm trained on emoji knows when you're being sarcastic on Twitter. MIT Technology Review.

[39] Rahwan, I. (2019). What can we learn from emojis? – MIT Media Lab. MIT Media Lab.

[40] El Ali, A., Heuten, W., Wallbaum, T., CJ Boll, S. and Merlin, M. (2017). Face2Emoji: Using Facial Emotional Expressions to Filter Emojis.

[41] Shah, R. and Kwatra, V. (2016). All Smiles: Automatic Photo Enhancement by Facial Expression Analysis. Google Research, Mountain View.

[42] Vemulapalli, R. and Agarwala, A. (2019). A Compact Embedding for Facial Expression Similarity. Google AI.

[43] Marinoiu, E., Zanfir, M., Olaru, V. and Sminchisescu, C. (2016). 3D Human Sensing, Action and Emotion Recognition in Robot Assisted Therapy of Children with Autism. Department of Mathematics, Faculty of Engineering, Lund University, Institute of Mathematics of the Romanian Academy.

[44] Shafran, I. and Mohri, M. A Comparison of classifiers for detecting emotion from speech. Center for Language and Speech Processing. Johns Hopkins University, Courant Institute of Mathematical Sciences New York University.

[45] Hickson, S., Dufour, N., Sud, A., Kwatra, V. and Essa, I. (2017). Eyemotion: Classifying facial expressions in VR using eye-tracking cameras. Georgia Institute of Technology, Google.

[46] Jain, V., Aggarwal, P., Kumar, T. and Taneja, V. (2017). Emotion Detection from Facial Expression using Support Vector Machine. International Journal of Computer Applications.

[47] Apprendimento Automatico.it. (2019). Emotions Detection Via Facial Expressions with python & OpenCV - Machine Learning & Cognitive.

[48] Fratesi, A. (2015). Automated Real Time Emotion Recognition using Facial Expression Analysis. Carleton University Ottawa, Ontario.

[49] Neves, António & Lopes, Daniel. (2016). A practical study about the Google Vision API.

[50] Smashing Magazine. (2019). Powerful Image Analysis with Google Cloud Vision and Python — Smashing Magazine.

[51] Chen, S. and Chen, Y. (2019). A Content-Based Image Retrieval Method Based on the Google Cloud Vision API and WordNet.

[52] Laurens van der Maaten. (2019). t-SNE. [online] Available at: https://lvdmaaten.github.io/tsne/ [Accessed 10 Jan. 2019].

[53] Titcomb, J. (2016). Apple buys company that scans your face to read emotions. [online] Telegraph.co.uk. Available at: https://www.telegraph.co.uk/technology/apple/12088601/Apple-buys-company-that-scans-your-face-to-read-emotions.html [Accessed 14 Jan. 2019].

[54] Michelle Yik, Sherri C. Widen & James A. Russell (2013): The within-subjects design inthe study of facial expressions, Cognition & Emotion, DOI:10.1080/02699931.2013.763769.

[55] Ngo, Hau & Rakvic, Ryan & Broussard, Randy & Ives, Robert. (2009). An FPGA-based design of a modular approach for integral images in a real-time face detection system. 10.1117/12.820248.