FACIAL RECOGNITION AND EMOTION DETECTION ALGORITHM OVERVIEW

[F.R.E.D.A. Overview]

Team Members:

Vinit R Iyer	19BSR06024
R Bhargavi Prakalya	19BSR06016
Anushka Kumari	19BSR06003
Rakhi Kumari	19BSR06036

Faculty Advisors:

Dr. Ramchand V
Mr. Sampangi Rama Reddy B R

Facial Recognition and Emotion Detection Algorithm Overview (FREDA Overview)

Vinit R Iyer, R Bhargavi Prakalya, Anushka Kumari, Rakhi Kumari

Abstract

Facial Recognition software has been widely used for the past decades and over the years, emotion detection has also been an active part of the research and building both facial recognition and emotion detection software in the same system has been crucial. It has great importance in the fields of computer vision and machine learning. This paper gives an overview of current algorithms and techniques of Facial Recognition and Emotion Detection Systems, Image Mining, Facial feature extraction, the advantages and disadvantages of Facial Recognition, and the current progress made in this field. This paper explores a few of the algorithms used in Facial Recognition and Emotion Detection in detail such as Eigen Based, Neural Networks, Local Binary Point(LBP), etc. for Facial Recognition and Support Vector Machine (SVM), Random Forests (RF), etc. for Emotion Detection.

Keywords – Facial Recognition, Emotion Detection, Algorithms, Image mining, Pre-processing, Eigen based, Neural Networks, Local Binary Points, Random Forests, SVM, LBP.

Introduction

Facial recognition is a method of identification or confirmation of a person's identity using their face as a reference [1]. An Emotion detection or a facial expression algorithm is one which is able to detect the emotion a person is experiencing based on the different displacement of the muscles of the face caused due to certain emotions. Facial and expression recognition is a category of Biometric security which is quite similar to other forms of biometric security such as finger prints, voice recognition, retina scanner, etc. It is commonly used in the Security sector but has now paved a path into various platforms.

Having set foot in the 21st century, we have successfully bridged the physical and virtual world, we have started viewing and analyzing the world in the form of data to a certain extent. Data are now being transmitted seamlessly and is giving rise to various technologies such as Artificial Intelligence, Augmented/Virtual Reality, Internet of Things, Cybersecurity and many more. One component of today's research area is known as Human-Computer Interaction which is the base for seeking an efficient communication between humans and machines. Currently, the channels of communication of humans with machines are the auditory channel (carrying the speech and vocal intonations) and visual channel (facial expressions and body movements). Facial expressions in such cases are of utmost importance as facial expressions are responses to a person's internal emotional state or intention.

Since the establishment of facial expression analysis as a research field in 1872 [2], facial recognition has been an active research topic across various fields such as Computer

Science, Biology, Psychology, etc. Especially in the field of Computer Science, Facial recognition, Image recognition, etc. are being utilized in an extensive range of applications such as Human-Computer Interaction, Biometric identification, Security, Intensive Care Monitoring in ICUs, Aerial image analysis, etc.

Currently, facial recognition is possible through partially covered faces too. At the moment of COVID-19 pandemic, wearing masks has become compulsory and facial recognition softwares failed to identify people wearing masks. A Chinese AI Company SenseTime pioneered a real world implementation of the Facial recognition technology which crosses the barrier of partial face coverings. It detects a total of 240 features from the human face majorly concentrated near the eye area. There was also a recent publication which shed light on a 90% identification success rate with only the bottom half of the face visible by analysts from the University of Bradford headed by Professor Hassan Ugail [11]. Apple has been working on and has made significant advances in facial recognition using sub-epidermal imaging technology. As facial recognition may fail to accurately identify and differentiate between twins and siblings, the current research conducted by Apple would create a milestone in Facial recognition and Emotion detection analysis considering each individual has a unique pattern of blood vessels underneath the skin of the face.

Psychological Background of Facial Recognition

The face plays an important role in social communication. Along with verbal communication, body language and facial expressions are things which affect the flow of the social interaction. In 1978, Ekman introduced the system for measuring facial expressions called Facial Action Coding System (FACS). The analysis of the relations between muscle contractions and changes in the face appearance caused due to fluctuating emotions are what gave rise to FACS. The expression analysis based on FACS made use of Action units, a decomposed set of the observed expression. There are 46 Action Units (AU) that represent change in facial expressions [3]. Action units are highly descriptive in nature but do not provide any information on the reason of the facial expression change. According to Ekman, there are six different emotions which are universal for all people of different ethnic places and cultures.

System Requirements

The system that is being designed and in progress is called a Facial Recognition and Emotion Detection Algorithm (FREDA). The FREDA system is composed of 4 main elements: Image Mining, Face recognition, Facial feature extraction, Emotion Recognition. Different methods are proposed for each of the elements, an overview for each is mentioned in this report. More in-depth research and related work would be conducted in further research.

The goal of FREDA system is to imitate human visual system in the most similar way. This requires a very efficient image/video analysis technique along with a well suited feature vector used in the Machine Learning Algorithm. A well suited data mining technique needs to be utilized which would reduce the noise in the data collected. FREDA being a fully automated system, needs to be effortless and efficient with limited human requirements. It is also preferred if the system is real time in nature. The system should be designed to avoid limitations on body and head movements which would be an important source of information about displayed emotions. The constraints of facial hair, accessories, cosmetics, etc. should be reduced to

minimum for minimizing the data pollution. These ocular problems are a challenge to the FREDA and needs to be taken into account [3].

Another important feature of FREDA is the user and environment independence wherein any user can be allowed to work on the system despite the skin colour, race, religion, etc. Environment dependence corresponds to the handling of complex background and varieties in lighting conditions, etc.

Steps involved in Facial Recognition and Emotion Detection

As per various literature surveys, it is found that implementing Facial Expression recognition is basically done in five steps [9], namely:

- **Image Mining**
  The data, or in this case images are gathered from the internet using data mining techniques such as object recognition, image indexing and retrieval, image classification and clustering, association rules mining, decision trees, and neural networks [10].
- **Pre-processing**
  The data collected from data mining techniques such as Deep Convolutional Neural Networks (DCNN), Haar Cascade Classifier, etc. is vast and filled with noise or garbage value. Thus most of the pre-processing steps implemented are mostly [9]:
  - Noise reduction
  - Image conversion to grayscale
  - Pixel brightening
  - Geometric transformation
- **Face Recognition**
  Facial recognition is a computer technology which determines whether the image is that of a human or not. In this step, faces are first located using a set of landmark points known as “face localization” or “face detection”. If faces are detected, then these are geometrically normalized to match the template image in a process termed “face registration”.
- **Facial Feature Extraction**
  Facial feature extraction refers to the process of locating key identification points or specific regions, mainly the eyebrows, eyes, nose and the lips.
- **Emotion Classification**
  Emotions are mainly divided into six basic emotions, namely: Anger, Disgust, Fear, Happiness, Sadness and Surprise. The process of emotion classification refers to process of allotting the images and analyzing the features and matching them to the corresponding basic emotion. Surprisingly, there are several thousands of human emotions according to the “Atlas of Emotions” by Professor Paul Ekman, a professor emeritus at the University of California, San Francisco [2].

Further research on each of the 5 steps has been done and different algorithms perused for Facial Recognition and Emotion detection, techniques for Facial Feature Extraction, Image mining and Pre-processing are explored and the most suitable are explained in detail.

Image Mining

The main purpose of Image Mining is to extract relations and patterns which are not directly stored in the database. It is an approach based on data mining, artificial intelligence, machine learning, image extraction, image processing. [17] It mainly deals with the extraction of data which are not directly available in our databases. Till date, Image mining is in its early phase. The major problem we face today in image mining research is that most of the researchers believe that image mining is just an extension of data mining or just a different name for pattern recognition [16]. However, it is different from low-level computer vision and image processing techniques as the main target in Image mining is to extract data from large-scale image data. In image mining, the objective is to locate image samples which are important for the given set of data [16].

An image database that contains unprocessed data is not fit to be directly used for image mining. Therefore, it has to be first processed so that it can be used in high-level mining modules. If an image mining system supports user interaction during extracting the patterns and knowledge from the collection of huge image database, then it is considered as a good technique. The two kinds of frameworks of image mining are:

- ***Function driven framework*:** It mainly focuses on various modules component and their functionalities. Most of the image mining system architectures existing today come under the function-driven image mining framework. These descriptions are particularly application-oriented and the framework was assembled according to the module functionality.
- ***Information driven framework*:** It provides a hierarchical structure of levels and the data needed into all the levels [16]. the purpose of The function-driven framework is to organize and clarify the different roles and tasks to be performed in image mining. Although, it fails to highlight the different levels of information representation necessary for image data. Hence, an information-driven framework was then proposed which highlighted the role of information at various levels of representation. The framework, distinguishes four levels of information.

The techniques frequently used in Image mining include object recognition, image indexing and retrieval, image classification and clustering, association rules mining, and neural network The techniques are classified on five levels of information and the associated image or data mining operations. Following are the details of some of the most important techniques:

- **Object Recognition**

  Object recognition has been a main focus for research in the field of image processing. An object recognition system finds objects in the real world from an image using object models which are known a priori. This is one of the major function of image mining. Generally, an object recognition module comprises of four components, known as, model database which consists of all the models known by the system and the model contains important features that describe objects, feature detector wherein the detected image primitive features in the Pixel Level are used to help the hypothesizer to assign likelihood to the objects in the image, hypothesizer and hypothesis verifier which uses the models to verify the hypothesis and refine the object likelihood.

- **Image Retrieval**

  Image mining requires that images be retrieved according to some requirement specifications. The classification is based on the increasing level of complexity:

  - ***Level 1:***

    It comprises image retrieval by primitive features such as color, texture, shape or the spatial location of image elements.

  - ***Level 2:***

    It comprises image retrieval by derived or logical features like objects of a given type or individual objects or persons.

  - ***Level 3*:**

    It comprises image retrieval by abstract attributes, involving a significant amount of high-level reasoning about the meaning or purpose of the objects or scenes depicted.

- **Image Indexing**

  Image mining systems require a fast and efficient mechanism for the extraction of image data. Currently, the retrieval of most image retrieval system is, by nature, similarity-based retrieval. In this case, indexing has to be carried out in the similarity space. One promising approach is to first perform dimension reduction and then use appropriate multi-dimensional indexing techniques that support Non-Euclidean similarity measures. Indexing techniques used range from standard methods such as signature file access method and inverted file access method, to multi-dimensional methods such as KD-B tree [19], R-tree [20], SR-tree [21], TV-tree [22], etc.

- **Image Classification and Image Clustering**

  Image classification (also known as, Supervised classification) is a technique where we are given a set of labelled images which are further used to label a new image by machine learning techniques. In image clustering, the main issue is to form sets of data (labelled) which is relevant from a given collection of un-labelled images. This step is generally done in the beginning phase of the Image mining process. The main purpose is to acquire content information the users are interested in, from the image group label associated with the image. The classification module in the mining system is usually called classifier. There are mainly two major types of classifiers, the parametric classifier and non-parametric classifier. Feature attributes that have received most attention for clustering are color, texture and shape. Once the images have been clustered, a domain expert is needed to examine the images of each cluster to label the abstract concepts denoted by the cluster [16].

- **Impact of Neural Networks in Image Mining**

  Neural networks are the computing systems which are mainly designed to recognize patterns. Their architecture is influenced by the human brain structure, hence the name. They consist of three types of layers: input, hidden layers, and output. The input layer receives a signal, the hidden layer processes it, and the output layer makes a decision or a forecast about the input data [18]. Each network layer consists of interconnected *nodes (artificial neurons*) that do the computation. Neural networks are fault tolerant and are good at pattern recognition and trend prediction. In the case of

limited knowledge, artificial neutral network algorithms are frequently used to construct a model of the data [16].

Preprocessing

Pre-processing is one of the most important steps which has a huge impact on the overall system. The data once collected must be processed thoroughly in bids of reducing the amount of storage space substantially while making the data more practical and easier to implement in the algorithms. The aim of preprocessing is the improvement of the image data by suppressing the unwilling distortions or by enhancing the image features crucial for further image processing. Multiple techniques can be implemented such as geometric transformations wherein rotation, scaling, translation, etc. are some important techniques in geometric transformations.

Image preprocessing techniques are mainly classified into 4 distinct categories based on the size of the pixel neighborhood that is utilized for the calculation of new pixel brightness. A few techniques are elaborated further:

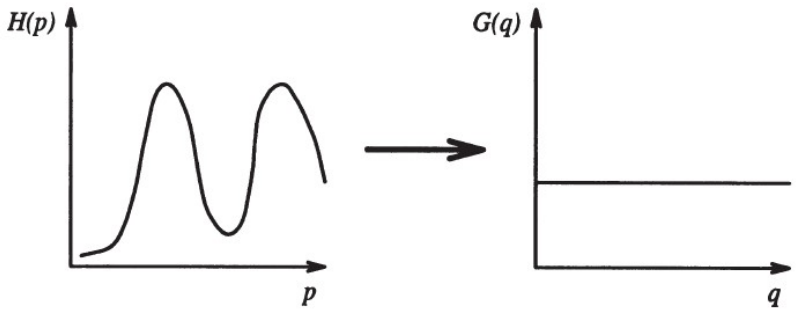
- **Pixel Brightness Transformations**

  A brightness transformation modifies the pixel brightness wherein the transformation depends on the properties of the pixel itself. Pixel brightness transformation are divided into 2 classes, brightness corrections which modifies the pixel brightness based on its original brightness and position of the image and grey scale transformation focuses on changing brightness regardless if the position of the image. The position dependent brightness correction can be represented in the following way:

$$f(i,j) = e(i,j)\, g(i,j)$$

  Where A multiplicative error coefficient $e(i,j)$ describes the change from the ideal identity transfer function; assume that $g(i,j)$ is the original un-degraded image (or desired or true image) and $f(i,j)$ is the image containing degradation. This method works on the assumption that linearity of the transformation.

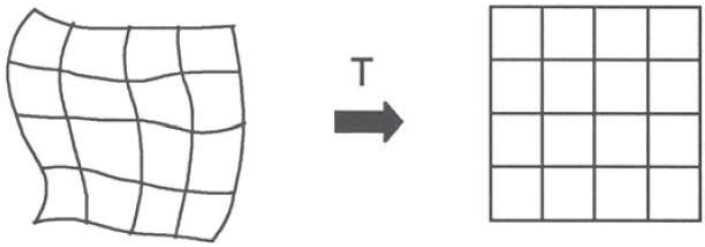
  A grey scale transformation for contrast enhancement would utilize a histogram equalization technique. A glimpse of the histogram equalization technique is given below.



Histogram equalization.

- **Geometric Transformations**

  Most commonly used in computer graphics and often used in image analysis. Geometric transformation eliminates the geometric distortion when an image is captured. An example for such a type of transformation is given below:



Geometric transform on a plane.

A geometric transformation consists of 2 basic steps: Pixel coordinate transformation followed by determination of brightness value through comparison of transformed point and the point in a digital raster. Coordinates of a point in the output image is usually approximated by the following polynomial equation.

$$x' = \sum_{r=0}^{m} \sum_{k=0}^{m-r} a_{rk} x^r y^k$$

$$y' = \sum_{r=0}^{m} \sum_{k=0}^{m-r} b_{rk} x^r y^k$$

- **Local Pre-processing**

  In this method a small neighbourhood of a pixel in an input image to produce a new brightness value for the output image. Local pre-processing methods can be reduced into 2 groups based on the aim of the processing. One method is based on smoothing which aims to suppress noise or other small fluctuations in the image, unfortunately it blurs the image and sharp features. The other group are called gradient operators which are based on local derivatives of the image function. They have a similar effect to suppressing lower frequencies in the Fourier transform domain but if this method is applied, the noise level increases. The contribution of the pixels in the Neighbourhood O is weighted by coefficients h using the following formula:

$$f(i,j) = \sum_{(m,n)} \sum_{\in O} h(i-m, j-n)\, g(m,n)$$

Facial Feature Extraction

With the emergence and improvement in the field of information technology, the identification and authentication among the human population has increased. Facial Feature Extraction stands up to be the most reliable, convenient and secure method for authentication. FFE is a multi-disciplinary application that has been developed with the increasing demand for the better authentication and information security. The Feature extraction reduces the redundant data and from the data set to build the model with less machine's effort and increases

the speed of learning and generalization of the machine learning process. Feature based approaches are built upon local facial features extracted from local components such as eyes, nose, mouth, etc. and local segmented regions. A few methods of facial feature extraction are also elaborated further:

- **Gabor Wavelets based Method**

  Gabor wavelets transformation is a dynamic joint time & frequency tool used for image analysis closely related to Gabor filter. The most important function of the wavelets is that it reduces the product of its standard deviation in reference to Time and frequency domain. Gabor wavelets have been used for years in numerous feature extraction algorithms. Using 2-D Fourier transformations on a set of Gabor kernels (also known as Gabor filters) limited by different orientations (usually 8) and scales (usually 5) and an input image, the method generates complex coefficients called Gabor wavelets components, which can be expressed by real and imaginary parts, or alternatively, by magnitude and phase parts. These components are representations of the given image and can be used for facial feature extraction.

- **LPQ Based Methods**

  In the past, Local Phase Quantization (LPQ), a definition of light tolerance, has been identified and studied for Facial Recognition (FR). By working with blurry images on the texture separation problem, LPQ stands out better than Gabor Wavelet-based methods and LBP methods. Having experience with the issue of LBP (which was originally developed to separate the texture but soon gained a better reputation as a feature of the feature in FR), many researchers have re-evaluated the use of FR LPQ. Although portrayed as insensitive, the LPQ has reported impressive reactions when dealing with blurred facial expressions. Based on the statistical dimension factor of the image phase spectrum on the background of frequencies, the LPQ operator in the image pixel is done using Short-term Fourier transform (STFT) over a $M \times M$ size window, the center of which is the image itself, with four scalar frequencies. The four imaginary parts and the four real parts are white based on the parameter ρ before they are measured in binary to obtain the given pixel LPQ label.

- **Other Methods**

  Other local definitions, such as the Scale Invariant Feature Transform (SIFT) and Histograms of Oriented Gradients (HOG), are often used in many real-world systems because of their effective calculation, partial closure resistance, and lack of sensitivity to viewing changes. Although SIFT and HOG are clearly shown to be the two best forms of edge code or location information, there are not many of their contributions to building a robust FR system. According to, the performance of HOG elements in the FERET database is much worse than that of LBP and Gabor wavelets. In short, this evidence means that SIFT and HOG cannot pave the way to the removal of a solid facial feature as expected from this thesis. Since face pictures taken under the context of video viewing have low resolution while gallery images are generally of high quality, it is good to have test images with better resolution. To do this, high-resolution techniques are used to produce high-resolution images before the feature is released in the hope of improving system accuracy.

  Steps involved Feature extraction methods in FR systems can be categorized into three types:

- Standard methods - these methods are based on the edges, lines and curves in the image input.
- Model-based methods - these methods are used to obtain real facial features such as eyes, nose, mouth.
- Structural matching methods - these methods take the geometric barriers to the elements considered. Based on psychological research on how people use complete and spatial features to identify faces.

Facial Recognition

Facial recognition is a method of identification or confirmation of a person's identity using their face as a reference [1]. Facial recognition has been the active area of research for the past decades. It has been a part of larger discipline of computer vision [23]. It is a category of biometric security. There are other forms of biometric security as well like fingerprint recognition, iris or eye retina recognition, voice recognition. But face recognition is preferred over the other biometrics because it is easy to use and implement. Multiple algorithms can be used for facial recognition, a few of which are:

- **Eigen Based**

  The aforementioned characteristic features (eyebrows, eyes, nose and lips) serve as the Eigenface [24] in the facial recognition domain. One way to extract these Eigenfaces is through the usage of a mathematical tool called Principle Component Analysis (PCA). Using the PCA, one can transform the original image from the training set into a set of Eigenfaces. One important feature is that it can create Eigenfaces from images and it can also create images from a set of complimentary Eigenfaces which has been built in some kind of weighted sum of all Eigenfaces. If all Eigenfaces from the original image extracted, is utilized, one can create the exact same source image. But in real life scenarios, a part of the original data is trimmed out and only a part of the Eigenfaces can be utilized which will give an approximation of the image then. The approximation can be done to a higher accuracy by choosing the most prominent features of the original data.

  It is possible to not only extract the face from Eigenfaces given a set of weights, but one can extract the weights based on the Eigenfaces and thus reconstruct the face to be recognized. This allows the confirmation of the image in question to be a face by comparing the weights obtained with the databank. If the values are very different to the weights of Human weights, then the image may not be a human face. A simple algorithm utilizing Eigenfaces would be the following:

  The original images of the training set are converted into a set of Eigenfaces E. The weights are calculated for each image in the training set and stored in set W. In the usage phase, upon observing an unknown image X, weights are calculated for the image and stored in vector WX. WX is then compared with the weights obtained from the training dataset to ascertain whether the image being scanned is a face or not. An eigenvector of a matrix is a vector such that, if multiplied with the matrix, the result is always an integer multiple of that vector. This integer value is the corresponding eigenvalue of the eigenvector. This relationship can be described by the equation:

$$M * \nu = \lambda * \nu$$

Where ν is an eigenvector of the matrix M and λ is the corresponding eigenvalue. A method for recognizing whether a given image is a human face or not is by utilizing Euclidean distance measures. Euclidean distance between two feature would be represented as:

$$d(xi, xj) = \sqrt{\Sigma}(ar(xi) - ar(xj))2nr = 1$$

Where $ar(x)$ denotes the value of the rth feature of an image X [24].

- **Neural Network Based**

  First, the facial recognition algorithm was run and the output for the facial recognition was found. Then the extracted features from the images are used as the input for the emotion/expression detection using a classification method to distinguish between emotions. The classifier should be able to recognize emotions irrespective of the gender, age, ethnic groups, lighting conditions, exposure, shadows, birthmarks, beard or glasses. There have been many researches in the field of Neural networks such as Gargesha and Luchi's proposed Multi-Layer Perceptrons and Radial basis function Networks. This method utilizes geometric coordinates of the facial characteristics, Euclidean distances for the contour points, etc. and the data collected input in the neural network fiving an accuracy of up to 73% with the JAFFE database [25]. An example algorithm for face detection is the following [26]:

  According to the research, feature value is calculated by subtracting sums of pixels covered by white rectangle from sum of pixels under grey rectangle. Input image is then transformed into an integral image in which each pixel is a sum of all pixels above and to the left.

$$ii(x, y) = \Sigma i(x', y')x' \leq x, y' \leq y$$

  Where $ii(x, y) = integral\ image\ i(x, y) = input\ image$. Another method is Lisetti and Rumelharts' neural network which classifies emotions based on the signaled emotions and the level of expressiveness. The neural network consists of a hidden layer which is connected to each part of the face and output unit thus giving a much more accurate result as seen by their results.

  Ensemble neural network had also been researched by Padgett wherein each neural network of the ensemble consists of a hidden layer of 10 nodes. Each of the neural network is trained independently and the output of each neural network is represented in a percentage value of each emotion.

- **Fisher Faced**

  Fisher linear discriminant analysis was developed by Robert Fisher in 1936 for the classification of flowers. It is a technique in pattern recognition that performs a class dimensionality reduction. The intention of this algorithm is that similar classes are clustered together, while different classes will be scattered as far away as each other. It minimizes the spread of variance to other classes. The Fisher-face algorithm uses inner class information for face classification. It can use multiple faces of a person to establish in-class variation in order to maximize class separation.

- **Linear Discrete Analysis**

  LDA (Linear discriminate analysis) is a statistical technique which is used to classify objects into mutually exclusive groups based on a set of unique features. The features are the observed faces and the groups can be classified as recognized and unrecognized. Linear means that groups are separable by a linear combination of features, like if there are only two features, then the separation between the object groups becomes a line and if there are three features, the separator is a plane and for more than three features, the separator would be a hyper plane. Similar to PCA, LDA is also used as a dimensionality reduction technique to project a dataset onto a lower-dimensional space [23].

- **Local Binary Point**

  Local Binary Patterns (LBP), a simple, effective texture based algorithm operator in computer vision that marks pixels in an image by setting each pixel's neighborhood threshold and treating the result as a binary number. Texture is an important characteristic of images. LBP can be used to know information about physical properties of objects like smoothness, roughness, or difference in surface reflectance such as color [23, 27]. The LBP operator processes an image as a composition of small patterns whose histogram reveals information about the distribution of edges and other local features. The term "operator" refers to a mapping function that can transform an image from one form to another [23]. The LBP algorithm creates histograms for each image that is labeled and classified. Each histogram represents each image from the training set. This way, the actual recognition process implies comparing histograms of any two images [28].

Emotion Detection

Emotion detection or emotion recognition is a classification task and is based on Machine Learning theory. The output of the Facial Feature Extraction step is utilized as an input for this step. This step requires supervised training and hence requires labelled date for the training purpose. The most commonly used Facial Expressions Classification method is that of Action Units or also known as FACS (Facial Action Coding System) in terms of the basic human emotions [3]. A variety of machine learning algorithms can be utilized for emotion detection such as K-Nearest neighbors, Artificial Neural Networks, Support Vector Machines, Hidden Markov Models, etc. The principle steps for utilizing these algorithms are divided into 3, namely, choosing an efficient feature set, an efficient machine learning algorithm and a diverse training dataset for the training part of the automation. There are multiple techniques that can be used for emotion detection in general. These include:

- **Speech**

  Emotion detection from speech is based on the extraction of features from acoustic signals. A numerical vector is created corresponding to points on an acoustic graph and machine learning algorithms work on these vectors to determine the emotion of the person through the speech. Multiple parameters such as pitch, Energy, Entropy, Zero Crossing Rate (ZCR), Spectral Spread, etc. are utilized. The number of parameters utilized can exceed well above 72 [14].

- **Facial Expression**

  The major content of this research is based on emotion detection based on recognizable faces. There are various techniques used for the detection of emotions through facial expressions such as: Principle Component Analysis (PCA) where the dimensionality of the image is reduced and an appropriate facial indexing is conducted. It is also known as Eigen face Approach. Fisher's Linear Discriminant (FLD) is another technique which is used which reduces scattering in an image dataset and has better performance than PCA. [13]

- **Text**

  For textual data there is Keywords based approaches which use synonyms and antonyms to determine word sentiments or to predict the semantic orientation of adjective based on a set of seed opinion words [12]. There is also a categorical classification used as an approach via Vector Space Models wherein matrix of co-occurrence frequency vectors is used to represent the dataset in a dimensional format. In this approach words are arranged in rows and sentences/paragraphs/documents in columns thus making a relationship between the rows and columns thus allowing the Vector Space Model to do its work.

- **Body gestures**

  Little attention has been given to body gestures and affective body postures even though body gestures are an important part of portraying emotional states to the observer. The studies in psychology shows that the combined visual impact of facial expressions with body gestures are the most informative [15].

There are multiple algorithms than can be utilized as an algorithm for emotion detection analysis, namely:

- **Support Vector Machine (SVM)**

  Support Vector Machine (SVM) are one of the most powerful classification algorithms wherein an optimal hyper plane is found which divides 2 classes accurately with a maximum margin to avoid any overlapping between two classes. In the case of Facial expressions, a multi-class SVM is used to denote each of the six basic emotions [29].

- **Random Forests**

  Random Forest's (RF) sometimes prove to be more dynamic than SVM in certain cases. Random forests are based on decision trees, but instead of just one classifier, one utilizes more number of forests (classifiers) to decide the class (in this case emotion) of the target variable.

- **Neural Networks**

  A type of Neural Network with Back propagation algorithm was once used for the detection of emotions/facial expressions. Back propagation algorithm is used for training and testing the extracted features from the characteristic features of the face. The proposed Neural Network model consisted of an input layer, an output layer consisting of six nodes representing the six basic emotions and a hidden layer which consisted of a number of nodes equivalent to the number of extracted features [25].

Facial recognition and its diversified application

- **Increased Security:** On a governmental level, facial recognition has helped identify criminals and terrorists leading to their apprehension and the indirect saving of millions of lives.

- **Reduced crimes:** With the cameras being everywhere and ease of tracking criminals with the use of facial recognition systems, the rate of crime has decreased a lot.

- **Integration Capability:** Facial Expression Recognition is a technology which is integrable in various multi-disciplinary fields as it is easily integrated.

- **Faster Processing:** The process of facial recognition is quite fast and at most take a few seconds. In the digital era of cyber-attacks and advanced hacking tools, facial recognition enables quick and efficient verification of a person's identity.
- **Greater Convenience:** As the technology becomes more widespread, the applications related to the facial recognition will also spread in different domains. People might be able to use their face for payments saving time, reduce physical touch at the time of need such as COVID, etc.

Current shortfalls of Facial Recognition in the field

- **Surveillance:** Facial recognition requires a camera and utilizes data analytics and thus gives rise to the fear of the unknown of privacy breach and data leakage. Although Facial Expression recognition is a highly integrable and convenient technology, it does have its own set of risks thus requiring suitable measures to prevent mal-usage.

- **Scope of Error:** Facial Expression recognition is not free of errors, a slight change in camera angle, a slight change in appearance might throw facial recognition off unless a powerful algorithm is created to filter out the features used to deceive the Facial expression recognition algorithm. An example can be that during facial recognition, the tattoos, scars, etc. are removed if analyzed to be fake.

- **Breach of Privacy:** The database consists of pictures of individuals, which sometimes are present without the consent of the said person. A regulatory framework would have to be created to prevent the privacy and ethical abuses.

- **Massive Data storage:** Facial recognition relies heavily on machine learning technology which requires a massive database to learn and thus deliver accurate results. Such large datasets require robust data storage. Small and medium size companies may not have sufficient resources to store the data thus leading to curbing of the spread of the technology. One way to prevent this would be using cloud computing. A break through in cloud storage and making it cheaper at the same time would allow for small and medium size companies to hold large amounts of data with their financial capacity.

Potential Challenges faced by FREDA

Although the current technology of Facial Recognition and emotion detection is already mature and no longer a substantial problem, it is still a challenge for computers to make real time inferences. There are various challenges such as difference in head shapes for different ethnicities, facial deformities caused due to natural and unnatural causes, the background differences for the same image, lighting changes, facial filter usages such as from Instagram and snapchat, etc. Current repository of facial expression datasets such as CK+ [4], JAFFE [5], MMI [6], RaFD [7], Oulu-CASIA [8], etc. are posed by trained actors and fundamentally differ from naturally occurring facial expressions seen in real life scenarios. Usages of accessories such as glasses, hats, scarfs, masks, etc. are also another challenge faced by such systems. Another major problem is the sheer number of mixed emotions portrayed by humans, for example: disgust with anger, sadness with fear, happiness with sadness, etc.

Conclusion

The aim of this literature overview was to explore the area of Facial Expression Recognition. Covering the topics of different algorithms used for Facial recognition, Emotion detection, techniques used for image mining, pre-processing, etc. The aim of this report was successful as a clear understanding of various concepts was gained. The report also captures the current breakthroughs of Facial recognition.

In the future work, we would like to proceed with the technical report of a Facial Expression Recognition using the knowledge gained from this thesis and gain an understanding on the feasibility of the combination of multiple algorithms in each step of the process of Facial recognition and Emotion Detection. Another interesting areas of research would be a suitable algorithm to protect the facial recognition software from privacy breach and data leak. Another interesting area of research could be a Facial Recognition system which can successfully differentiate between fake features on a face and recreate the real face. Finally, pursuing research to minimize the disadvantages of the Facial Expression Recognition would be a broad pathway and yet the most required considering the progress of the technology today.

References

[1] Yu Miao, "A real time Facial Expression Recognition System using Deep Learning", submitted to University of Ottawa, 2018.

[2] https://www.paulekman.com/blog/introducing-atlas-emotions/

[3] Ewa Piatkowska, "Facial Expression Recognition System", Technical report submitted to DePaul University, 2010.

[4] Patrick Lucey, "The extended Cohn-Kanade dataset (ck+): A complete dataset for action unit and emotion specified expression", presented in IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshop, 2010.

[5] Michael Lyons, "Coding Facial Expressions with Gabor Wavelets", presented in Proceedings of the 3rd International Conference on Face & Gesture recognition, 1998.

[6] Maja Pantic, "Web-based database for Facial expression analysis", presented in Proceedings of the IEEE Conference on Multimedia and expo, 2005.

[7] Oliver Langner, "Presentation and validation of the Radboud Faces Database", in Cognition and emotion 24.8, 2010.

[8] Guoying Zhao, "Facial Expression Recognition from near infrared videos", in Image and Visual computing 29.9, 2011.

[9] Angana Mitra, Souvik Choudhury, Susmita Moitra and Mr. Amit Khan, "Image processing Facial Expression Recognition", submitted to RCC Institute of Information Technology, 2018.

[10] Ji Zhang, Wynne Hsu, Mong Li Lee, "Image Mining: Issues, Frameworks and Techniques", presented at the 2nd ACM SIGKDD International Workshop on Multimedia data mining, 2001, San Francisco, California, USA.

[11] Ali Elmahmudi, Hassan Ugail, "Deep face recognition using imperfect facial data", published in Future Generation Computer Systems, Volume 99, Pages 213-225, 2019.

[12] VV Ramalingam, A. Pandian, Abhijeet Jaiswal and Nikhar Bhatia, "Emotion detection from text", presented in National Conference on Mathematics Techniques and its Applications (NCMTA 18), 2018.

[13] Jyoti Rani, Kanwal Garg, "Emotion Detection Using Facial Expressions – A Review", published in International Journal of Advanced Research in Computer Science and Software Engineering, Volume 4, Issue 4, 2014.

[14] Mikel deVelasco, Raquel Justo, Josu Anton, Mikel Carrilero, M. Ines Torres, "Emotion Detection from Speech and Text", presented in IberSPEECH, Barcelona, Spain, 2018.

[15] H. Meeren, C. Heijnsbergen, and B. Gelder, "Rapid perceptual integration of facial expression and emotional body language", Proceedings of the National Academy of Sciences of USA, 2005.

[16] Ji Zhang," IMAGE MINING: ISSUES, FRAMEWORKS AND TECHNIQUES", submitted to Department of Computer Science, School of Computing.

[17] Wahiba Abdessaleam, "Image Mining Framework and Techniques: A Review", article in International Journal of Image Mining, 2015.

[18] https://www.altexsoft.com/blog/image-recognition-neural-networks-use-cases/

[19] J. T. Robinson. The K-D-B tree: A search structure for large multidimensional dynamic indexes. In Proceeding of the 1981 ACM SIGMOD Conference, pages 10-18, June 1981.

[20] A. Guttman. R-tree: a dynamic index structure for spatial searching. In Proc ACM SIGMOD, 1984

[21] N. Katayama and S. Satoh. The SR-tree: An index structure for high-dimensional nearest neighbor queries. In proceedings of the 1997 ACM SIGMOD

[22] K. Lin, H. V. Jagadish and C. Faloutsos. The TVtree: An index structure for high-dimensional data. The VLDB Journal, 3 (4): 517-542, 1994.

[23] Farshad Ghahramani, "Face Recognition: An Engineering Approach", submitted to San Jose University, 2015.
[24] Dimitri Pissarenko, "Eigenface-based facial recognition", 2002.
[25] Watshala Nayomi Widanagamaachchi, "Facial Emotion Recognition with a Neural Network Approach", submitted to University of Colombo, 2009.
[26] Paul Viola, Michael J. Jones, "Robust Real Time Face Detection", published in International Journal of Computer Vision 57(2), 137—154, 2004.
[27] C.H. Chen, L.F. Pau and P.S.P Wang. The handbook of pattern recognition and computer vision (2nd Edition). Singapore: World Scientific Publishing Co, 1998, pp. 207-248.
[28] https://recfaces.com/articles/facial-recognition-algorithmsemotions/
[29] Nitisha Raut, "Facial Emotion Recognition using Machine Learning", submitted to San Jose State University, 2018.