Journal of Information Systems & Information Technology (JISIT) *Vol. 5 No.1 2020*
ISSN: 2478-0677 *1-13*

Training Computers to Recognize Facial Expressions of Humans: A Machine Learning Approach

Ahamed Rameez Mohamed Nizzad and Kunaratnam Kunarakulan

Department of Computing, School of Computing, British College of Applied Studies
nizzad@bcas.lk, gunarakulan@gmail.com

Abstract. Humans are precise in facial expressions and recognizing them. Facial expression is a routine task of humans. Training Computers to recognize such humanly task as closed to humans can be useful in numerous ways such as offering services as per the mode of consumer or user and will enhance the phenomenon of human computer interaction. Environment and surroundings are the sources from where humans naturally learn the models of facial expressions and they have logical models in brain, and with the comparison of models, they are able to recognize the expressions of others. This research aimed to design and develop a robust facial expression recognition system by combining various techniques available in Computer Vision and Machine Learning. Hence, it was crucial to understand the human psychological aspects of facial expression in a precise manner. For the training and demonstration purposes, human volunteers were utilized along with facial expression dataset such as CK+ from internet. Local Binary Patterns (LBP) and Convolutional Neural Network (CNN) were occupied for feature extraction and optimization respectively. In addition, different kinds of AI based tools and techniques were employed to mimic the human ability of recognizing facial expressions. The outcome presents a novel approach to facial expression recognition with modified LBP and CNN. The outcome of the research suggests that further studies and optimization would lead to commercially viable solutions such as products or services as per the expression and emotions of the consumers. The researchers will continue to optimize the outcome to supersede humans in recognizing facial expressions in future.

Keywords: Facial Expression, Face Recognition, Machine Learning, Computer Vision, Convolutional Neural Networks (CNN)

1 Introduction

As computers become ever more abundant and multi-purpose, systems are being developed that attempt to recognize aspects of facial emotion related behaviors and feelings. It is a study on facial expression detection. It covers various facts and theories about facial expressions, psychology of emotions and other artificial intelligence-based algorithms and theories which are very important to implement the facial expression recognition system with the artificial intelligence technology.

In this technological world, Artificial Intelligence (AI) is one of the most advanced technologies in computing which is revolutionizing all industries. Computer vision is a sub set of AI where are many useful applications have been developed with the use of computer vision technologies such as object detection, face recognition and etc. The AI technology is about mimicking human abilities to computers. At present, computers are able to see real world objects, human faces, and they can even understand human voices and perform some actions according to instructions.

Received: 27th Mar, 2020
Revised and Accepted: 25th Jun, 2020

In computer vision, there are many technologies available such as face detection, face recognition, pattern recognition, image processing and etc. Therefore, with the help of those computer vision technologies, this research explores suitable algorithms to create a facial expression recognition system to track human emotions. Therefore, it is essential to study the psychological background of human facial expressions and underlying theories and principles.

A facial expression is one or more motions or positions of the muscles beneath the skin of the face. According to one set of controversial theories, these movements convey the emotional state of an individual to observers. Facial expressions are a form of nonverbal communication and humans have the ability to recognize facial expression naturally. There are some minor differences on facial expressions due to the cultural changes.

Alert Mehrabian studied and found that, nonverbal part is the most informative channel in social communication. This means verbal part contributes about 7% of the message, vocal is 34% and facial expression about 55% [1]. Paul Ekman found that, there are 6 basic universally accepted types of facial expressions. They are happiness, sadness, fear, surprise, anger, and disgust [2]. There are 7 types of basic face shapes available such as round, oval, square, oblong, heart, triangle and diamond and Face contains 46 action units (AC) which are responsible for muscles movements.

Fig. 1. Face Action Units

With the help of these psychological background of facial expression related evidence, this research attempts to mimic computers the ability to recognize facial expressions with the integration of machine learning techniques and algorithms.

2 Literature Review

2.1 Facial Expression in psychological and biological perspective

Face plays an important role in social communication. This is a ‘window’ to human character, reactions and ideas. Professor Alert Mehrabian, a psychologist introduced the communication theory that shown the nonverbal part is the most enlightening channel in social communication. Dr. Paul Ekman defined the basic six types of universally accepted facial expressions. Namely happiness, sadness, fear, surprise, anger, and disgust. The Facial Action Coding System (FACS) refers to a set of facial muscle movements that correspond to a displayed emotion. This theory called Facial Action Coding System was introduced and developed by analysis of the relations between muscles contraction and changes in the face appearance caused by them. There are 46 action units (AU) that represent changes in facial expression and 12 AUs connected with eye gaze direction and head orientation. Action Units are highly descriptive in terms of facial movements [3]. Cultural differences will make slightly minor different changes in Facial Expressions, according to the study by Professor Lisa Barrett, in the Northeastern University of America [4]. Further, the Friesen and Ekman discussed about how does the mind work for the facial expression in their book called book called Emotion Revealed [5].

2.2 Concept of Artificial Intelligence and subsets of AI

2.2.1 Artificial Intelligence

In the field of computer science, artificial intelligence (AI), sometimes called machine intelligence, is intelligence demonstrated by machines, in contrast to the natural intelligence displayed by humans and other animals. Computer science defines AI research as the study of "intelligent agents": any device that perceives its environment and takes actions that maximize its chance of successfully achieving its goals [6].

2.2.2 Computer Vision

Computer vision is an interdisciplinary scientific field that deals with how computers can be made to gain high-level understanding from digital images or videos. From the perspective of engineering, it seeks to automate tasks that the human visual system can do [7].

2.2.3 Pattern Recognition and Machine Learning

Pattern Recognition is the automated recognition of patterns and regularities in data. Pattern recognition is closely related to artificial intelligence and machine. Machine learning is seen as a subset of artificial intelligence. Machine learning algorithms build a mathematical model of sample data, known as "training data", in order to make predictions or decisions without being explicitly programmed to perform the task [8].

2.2.4 Face Detection

Face Detection is a computer technology being used in a variety of applications that identifies human faces in digital images [9]. Face detection also refers to the psychological process by which humans locate and attend to faces in a visual scene [10].

2.3 Algorithms and Technologies

The Viola–Jones face detection algorithm is the first object detection framework to provide competitive object detection rates in real-time proposed [11]. In machine learning, support-vector machines (SVMs, also support-vector networks) are supervised learning models with associated learning algorithms that analyze data used for classification and regression analysis [12]. A neural network is a network or circuit of neurons, or in a modern sense, an artificial neural network, composed of artificial neurons or nodes [13].

In deep learning, a convolutional neural network (CNN) is a class of deep neural networks, most commonly applied to analyzing visual imagery. CNNs use a variation of multilayer perceptron designed to require minimal preprocessing [14]. Local binary patterns (LBP) is a type of visual descriptor used for classification in computer vision. LBP is the particular case of the Texture Spectrum model proposed in 1990 [15]. Feature extraction is a dimensionality reduction process, where an initial set of raw variables is reduced to more manageable groups (features) for processing, while still accurately and completely describing the original data set (deepai.org).

2.4 Tools

Python is an interpreted high-level, general-purpose programming language. Created by Guido van Rossum and first released in 1991, Python has a design philosophy that emphasizes code readability, notably using significant whitespace. It provides constructs that enable clear programming on both small and large scales. TensorFlow is an open-source software library for dataflow programming across a range of tasks. It is a symbolic math library, and is also used for machine learning applications such as neural networks. OpenCV (Open source computer vision) is a library of programming functions mainly aimed at real-time computer vision. Haar-Cascade is Haar-like features are digital image features used in object recognition. They owe their name to their intuitive similarity with Haar wavelets and were used in the first real-time face detector. Anaconda is a free and open-source distribution of the Python and R programming languages for scientific computing (data science, machine learning applications, large-scale data processing, predictive analytics, etc.) that aims to simplify package management and deployment. Package versions are managed by the package management system conda ("Conda – Conda documentation".). The Anaconda distribution is used by over 6 million users and includes more than 1400 popular data-science packages suitable for Windows, Linux, and MacOS. (www.anaconda.com). The Canny edge detector is an edge detection operator that uses a multi-stage algorithm to detect a wide range of edges in images [16].

2.5 Facial Expression Datasets

A facial expression database is a collection of images or video clips with facial expressions of a range of emotions. Extended Cohn-Kanade Dataset is a facial expression data set which has neutral, sadness, surprise, happiness, fear, anger, contempt and disgust [17]. Japanese Female Facial Expressions (JAFFE) dataset [18] which has 213 images of 7 facial expressions such as 6 basic facial expressions and 1 natural expression.

3. Methodology

This research employs deductive approach to select the most efficient algorithm by comparing the existing theories, algorithms and tools & technologies available. This research requires the use of both mix method where qualitative and quantitative aspects are incorporated to reach the expected outcome. The psychological aspects are treated and analyzed by means of qualitative approach whereas machine learning involves quantitative approach such as statistical computation. Further the research relied on secondary sources (datasets) and some volunteer's facial expressions were captured to train and test in real time.

3.1 Conceptual Model

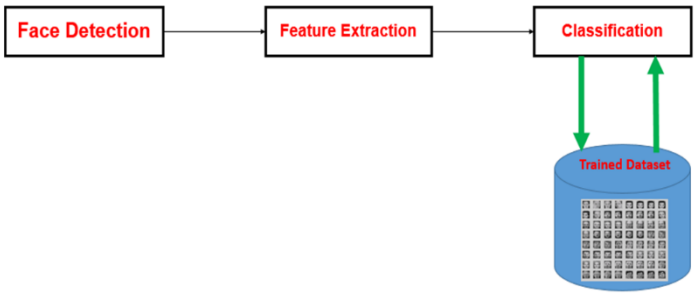

Fig. 2. Conceptual Model

The above diagram illustrates the steps required to recognize the facial expression. The major steps involves in this process are face detection, feature extraction and classification.

At the outset, the whole process is summarized into three stages. In every stage, there are several machine learning methodologies should be employed. Face detection stage uses some machine learning algorithms to target only the region of face from the video or image frame. Then it applies necessary filters to normalize the image to make it ready for feature extraction process. In the feature extraction part, there are a lot of deep learning methodologies have to be used. The first stage detects the face and applies some filters then throws the face into the feature extraction part. This feature extraction process targets some essential area of face which are useful for recognizing emotions. After the feature extraction process, facial parts will be converted as tensor floating point number to match against the dataset. Then the entire image will be converted as numbers array. After that, the whole array value which is produced by the feature extractor, sends to the classifier for classification purpose. The responsibility of the classifier is to match the extracted array with the trained tensor flow model. It acts like a gateway between tensorflow dataset and feature extractor. The classifier has the label emotions such as happy, sad, angry, surprise, fear and disgust.

3.2 Data Collection

The core data required for this study is facial expressions of humans. To increase the efficiency of the machine learning, the more facial expressions we train the more accurately it learns and so does the output. Therefore the researchers relied on following open source datasets available for the purpose of training the model and build a robust classifier. JAFFE Dataset, CK+ (Cohn-Kanade) Dataset, Over 1500 images with facial expressions from internet along

with facial expressions of some human volunteers who voluntarily agreed to capture their facial expressions where used for the purpose of conducting this research.

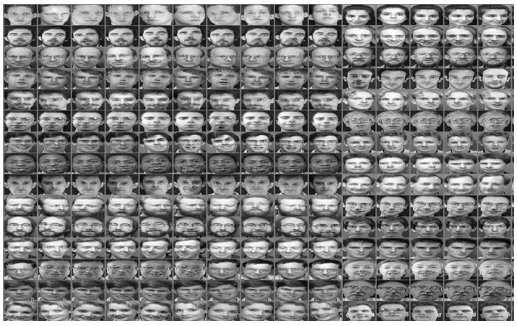

Fig. 3. Sample dataset with different facial expressions

4. Design and Development of the Model

4.1 Face Detection

As discussed above, at the beginning it is required to achieve the face detection when a video or image is input into the system. It is very important to detect the face in order to read the facial expressions. Without face, it would be impossible to read the emotions. From the input data, it might be Webcam, video file, image or external camera, the system has to target the face region according to extract the facial features. For achieving this, there available many algorithms and frameworks as per the literature review.

Face detection algorithms often begin by searching for human eyes. Eyes constitute what is known as a valley region and are one of the easiest features to detect. Once eyes are detected, the algorithm might then attempt to detect facial regions including eyebrows, the mouth, nose, nostrils, and the iris. Once the algorithm surmises that it has detected a facial region, it can then apply additional tests to validate whether it has, in fact, detected a face. Out of the algorithms such as Eigenfaces Algorithm, Viola and Jones Algorithm and Fisaher Face algorithm, the Viola–Jones object detection framework was chosen for the purpose of face detection. The properties of this algorithm are that training is slow, but detection is fast. This algorithm uses Haar basis feature filters, so it does not use multiplications along with Haar Classifier which is used to detect the object for which it has been trained for, from the source. The Haar Cascade is trained by superimposing the positive image over a set of negative images. The training is generally done on a server and on various stages. Better results are obtained by using high quality images and increasing the amount of stages for which the classifier is trained. In this study, Viola and Jones face detection algorithm with the use of "haarcascade_frontalface_default.xml" employed as this approach is much faster than other algorithms.

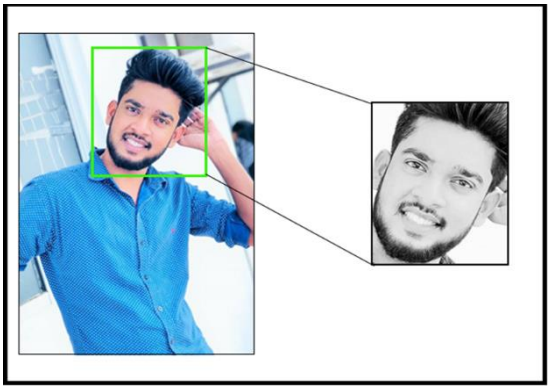

Fig. 4. Face Detection from an image

4.2 Feature Extraction

Once the face region has been targeted. This feature extraction step will extract the facial features such as eye, mouth and so on, with the use of geometric location method. This feature extraction part is the main function of the proposed system. In this section, there are many machine learning mechanisms would be used such as Convolutional Neural Network (CNN), Local Binary Patterns (LBP), landmark and etc. Before the feature extraction stage, the image normalization steps will be automatically proceeded to make the frames smooth and easy to detect the edges and wrinkles in them. Finally, it forms a facial model with the help of feature extracted data and the geometry of the face data. After that, it will send the formed facial model to classifier for the classification process.

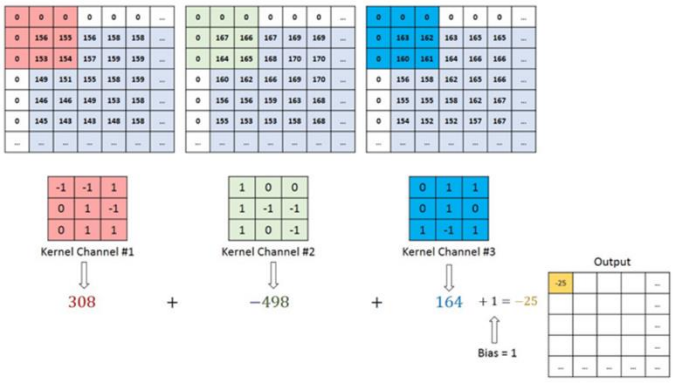

Fig. 5. Computation of feature extraction from detected face

In this step, it is required to Gaussion blur the detected face by blurring using guassian function in order to reduce the noise and detail in the image. Then the image is converted to grayscale image. Then using canny edge detector edges of the images are identified which use a multi-stage algorithm. Then locations of the facial landmark points around facial components are captured.

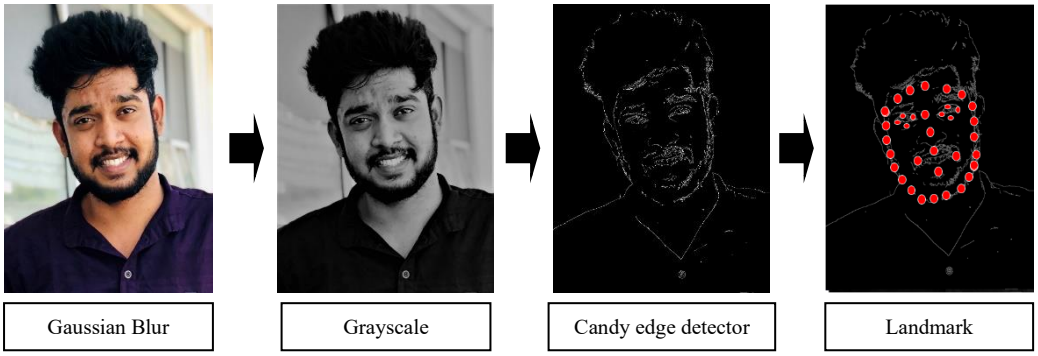

Fig. 6. The process of feature extraction

4.3 Classification

In machine learning, classification is a supervised learning approach in which the computer program learns from the data input given to it and then uses this learning to classify new observation. From the feature extraction part, it forms a model and check with the pre-trained data in a heavy brute-force method. If any matches found, it will return the name of the facial expressions. This is the final stage of the system. Basically, the feature extraction part, converts the input image as a numpy multi-dimensional array. After the conversion, further processing details will be carried out by the classifier. The classifier contains some labeled data. The classifier tries to match the input data with the trained dataset and after matching, it returns a probability value, based on that value the result is taken. The classifier already has trained data categories for each facial expressions such as happy, angry, fear, disgust, sad and surprise. Finally, which one gets high score will be the result.

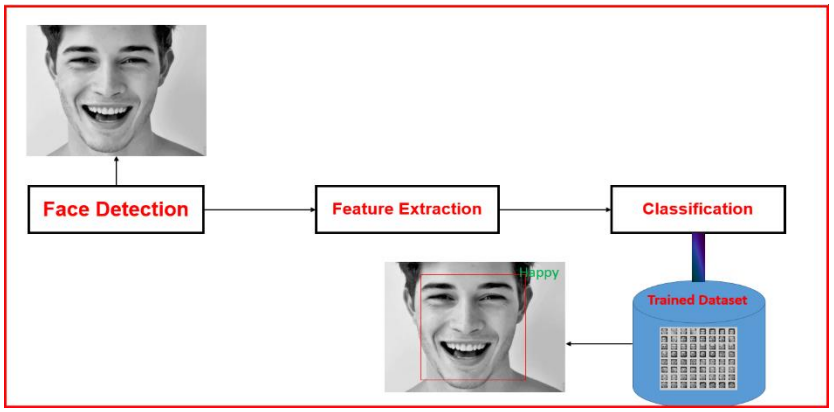

Fig. 7. Classification of Facial Expression

4.4 Training and Dataset.

In machine learning, the study and construction of algorithms that can learn from and make predictions on data is a common task. Such algorithms work by making data-driven predictions or decisions, through building a mathematical model from input data. The researchers utilized the datasets discussed above for training the model that has the ability to

learn the facial expressions from the images and videos. The process is just like teaching a child, this is dog or this is cat, like that system is trained to label the facial expression data. The training process also follows the same stages and procedures like classification expect after feature extraction the image is labeled as one of the facial expressions.

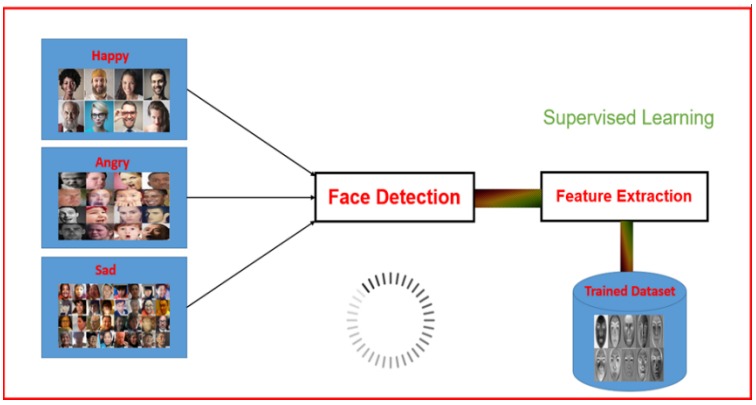

Fig. 8. Illustration of Training dataset classification

4.5 System Architecture Diagram

The above architecture explains the work flow of the facial expression recognition system. The process has been already explained in previous chapters. In addition to that, the effectiveness of the training depends on how much data is used to train. The more the data the effective is the model so does the classification.

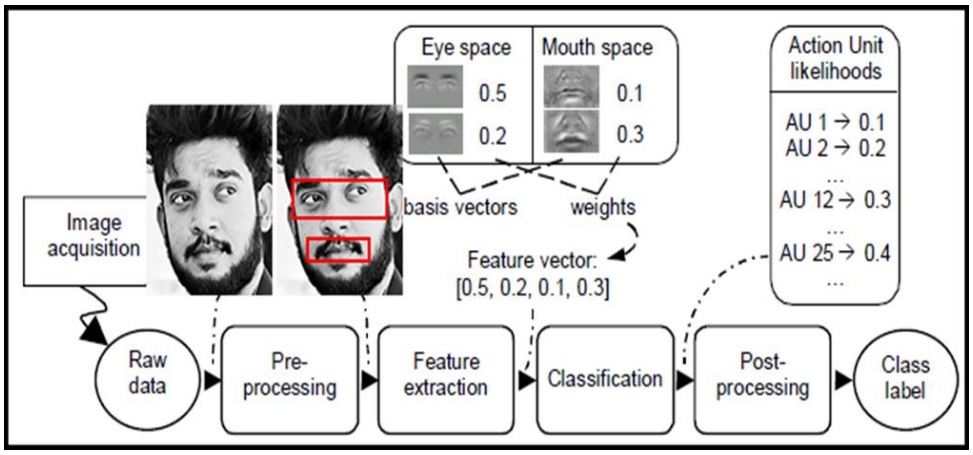

Fig. 9. System Architecture Diagram

5. Results and Discussion

The outcome of the implementation of the proposed model is able to predict the facial expression correctly.

Fig. 10. Output with labels as identified by the implemented system

The researchers found two algorithms for addressing facial expression. They are appearance based and geometric based. However, researchers relied on the first algorithm. The basic architecture of the proposed solution is face detection → feature extraction → Classification and the same approach was used to train the model where the extracted image is labeled and stored as tensors.

This research can be useful for many real world problems such as Surveillance, Human-robot interaction, Computer games' user experiences, Stress-monitoring, Human behavior analysis, Lie detection, Self-driving car like Tesla, Low-bandwidth videoconferencing and many. For instance, car manufacturers around are increasingly focusing on making cars more personal and safe to drive. Utilizing smarter car features, it makes sense for manufacturers to use AI / ML to help them better understand human emotions. Using facial detection, smart cars can alert the driver when he is feeling drowsy or lethargic or the engine can be stopped safely. Facial detection system can find subtle changes in facial micro-expressions that precedes drowsiness and send personalized alerts to the driver asking him to stop the car and go for a coffee break, change music or temperature etc.

In order to make a better facial expression recognition system, it is required to have a large dataset and need to be trained on Graphics Processing Unit (GPU) or high processing server computers. A general purpose computer would be slow for processing as the approach is brute force. It is highly recommended to use GPU for training the model yet i5 processor with minimum 4GB Ram are enough to run the application with the camera with moderate clarity. For the purpose of this study it took more than a day to train as the researchers didn't employ GPU However, had it been GPU it would have taken more or less 2 hours to train the model.

6. Recommendations

- This research utilized the appearance based algorithm but researchers recommend “geometric based method” though the approach is complicated but will result in more accuracy.
- It is recommended to viola and jones algorithm with the use of Haar-casecade face detection algorithm for face detection.
- For image processing, it is recommended to use Convolutional Neural Network (CNN) and also Local Binary Pattern (LBP) can be used.
- Preprocessing must be carried out before feature extraction and classification stages because, the preprocessing is a process which can be used to improve the performance of the system via removing noises from images.
- GPU computers are recommended for high processing and running the application.

7. Future Works

The limitation of the developed system is that the model had been trained utilizing limited dataset and literature suggests that cultural influences and perceived emotions have influence on how people are expressing their emotions via face. Therefore, the researchers will continue to work and remove this barrier by training as much as possible data and will extend the application by implementing functions such as playing games, music, food and etc triggered as per the mood of the user displayed by facial expression.

Acknowledgement

The researchers entirely acknowledge all the entities, respective volunteers who provided the dataset for this study, previous works of respective authors and institutions which helped the researchers to study the phenomenon. Researchers will responsibly acknowledge any sources left uncited though every steps have been taken to cite the respective works of others if brought to the attention of the researchers.

References

[1] A. Mehrabian and M. Wiener, "Decoding of inconsistent communications," Journal of Personality and Social Pyschology, vol. 6, no. 1, pp. 109-114, 1967.
[2] P. Ekman, "Facial Expression and Emotion," American Psychologist, vol. 48, no. 4, pp. 384-392, 1993.
[3] P. Ekman and W. Friesen, Facial Action Coding System: A Technique for the Measurement of Facial Movement, Palo Alto: Consulting Psychologists Press, 1978.
[4] L. F. Barrett, How Emotions are Made: The Secret Life of the Brain, Boston: Houghton Mifflin Harcourt, 2017.
[5] P. Ekman, Emotions Revealed, New York: Times Books , 2003.
[6] P. David, M. Alan and G. Randy, Computational Intelligence: A Logical Approach, New York: Oxford University Press, 1998.
[7] D. H. Ballard and C. M. Brown, Computer Vision, New Jersey: Prentice Hall, 1982.
[8] C. M. Bishop, Pattern Recognition and Machine Learning, Berlin: Springer, 2006.

[9] D. R. Frischholz, "The Face Detection Homepage," Face Recognition Homepage, 02 06 2003. [Online]. Available: https://facedetection.com/. [Accessed 08 12 2019].
[10] M. B. Lewis and H. D. Ellis, "How we detect a face: A survey of psychological evidence," International Journal of Imaging Systems and Technology, vol. 13, pp. 3-7, 2003.
[11] P. Viola and M. Jones, "Robust Real-time Object Detection," International Journal of Computer Vision, pp. 1-3, 2001.
[12] C. Cortes and V. N. Vapnik, "Support-vector networks," Machine Learning, vol. 20, no. 3, p. 273–297, 1995.
[13] J. J. Hopfield, "Neural networks and physical systems with emergent collective computational abilities," Proc Natl Acad Sci U S A, vol. 79, no. 8, p. 2554–2558, 1982.
[14] A. F. Vermeulen, Industrial Machine Learning: Using Artificial Intelligence as a Transformational Disruptor, New York: APress, 2019.
[15] D. He and L. Wang, "Texture Unit, Texture Spectrum, And Texture Analysis," IEEE Transactions on Geoscience and Remote Sensing, vol. 28, pp. 509-512, 1990.
[16] J. Canny, "A Computational Approach To Edge Detection," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 8, no. 6, pp. 679-698, 1986.
[17] P. Lucey, J. F. Cohn, T. Kanade, J. Saragih, Z. Ambadar and I. Matthews, "A complete facial expression dataset for action unit and emotion-specified expression," in IEEE Workshop on CVPR for Human Communicative Behavior Analysis, California, 2010.
[18] M. J. Lyons, S. Akamatsu, M. Kamachi and J. Gyoba, "Coding Facial Expressions with Gabor Wavelets," in 3rd IEEE International Conference on Automatic Face and Gesture Recognition, Nara, Japan, 1998.

Appendix

Modules Coded in Python

- Face_Detector.py – Responsible for detecting face and crops for training purpose.
- Frame_spliter.py – split face area from video clip and stores for training purpose.
- Sample_Genarator.py – This module extracts face area from webcam and applies image normalization techniques and store only the face region for training purpose.
- Tensorflow_Model_Convertor.py – This module is responsible for saving the system working data as tensor model for trouble shooting purpose.
- Run-(Feature Extractor).py – This is start point of the system. When this file runs, it calls all necessary modules and framework. And it also display the webcam to be alive and capture data. If any faces are in the video frame, it crop the face area and applies all important filters and sends to the classifier for classification purpose.
- Classifier.py – This is the main component of this system. This classifies the expressions.
- Train.py – This python module is used to train the model.