Advancements and recent trends in Emotion Recognition using facial image analysis and machine learning models

Tuhin Kundu
Dept. of Computer Science & Engineering
Jalpaiguri Government Engineering College
Jalpaiguri, West Bengal, India
Email: tk1910@cse.jgec.ac.in

Chandran Saravanan
Dept. of Computer Science & Engineering
National Institute of Technology
Durgapur, West Bengal, India
Email: dr.cs1973@gmail.com

***Abstract*—As the demand for systems with human computer interaction grows, automated systems with human gesture and emotion recognition capabilities are the need of the hour. Emotions are understood by textual, vocal, and verbal expression data. Facial imagery also provides a constructive option to interpret and analyse human emotional issues. This paper describes the recent advancements in methods and techniques used to gauge the five primary emotions or moods frequently captured on images containing the human face: surprise, happiness, disgust, normality, drowsiness, through automated machinery. Looking at the recent developments in facial expression recognition techniques, the focus is on *artificial neural networks* and *Support Vector Machine (SVM)* in emotion classification. The technique first analyses the information conveyed by the facial regions of the eye and mouth into a merged new image and using it as an input to a feed forward neural network trained by back propagation. The second method showcases the use of *Oriented Fast and Rotated (ORB)* on a single frame of imagery to extract texture information, and the classification is completed using *SVM*. The special case of drowsiness detection systems using facial imagery by pattern classification, as automated drowsiness detection promises to play a revolutionary role in preventing road fatalities due to lethargic symptoms in drivers is also discussed.**

***Index Terms*—Artificial Neural Networks, Emotion Recognition, Facial Analysis, Oriented Fast and Rotated BRIEF, Pattern Classification, Support Vector Machine**

I. Introduction

Emotions are human behavioural feelings that play a daily role in our day to day life in tasks such as decision making, learning, motivation, reasoning, attentiveness, planning and many more [1]. Humans are able to detect faces and interpret emotional expressions on them without any serious difficulty [2]. Considering the fast increasing interest in emotion recognition using applications [3], if automated systems successfully recognise human emotion, it opens the doors for automated analysis of human affective behaviour and attracts the attention of researchers from various disciplines such as psychology, linguistics, computer science, and other related disciplines [4] and opens up a big research field [1]. While various means of judging human emotion are available, analysis of facial images provide us with a near accurate means of contemplating human emotion and behaviour.

The implementation of a system that has automated facial expression or emotion analysis has three primary steps. The preliminary step includes, before facial imagery is analysed, a face must be detected in the scene or image sequence. This step includes multiple challenges such as face orientation, that is the angle between camera and face, and lighting, The second step includes *facial expression data extraction* which involves extraction of the prominent *facial features* such as eyes, nose, mouth, chin, and eyebrows. The final step which concludes the process is the *facial expression classification*. While there have been many studies to determine an effective way of classifying expressions on facial imagery, the *Facial Action Coding System (FACS)* [5] still is the most known study in this field till date [2].

Getting inspired from the structure of brain and biological nervous networks, *Artificial Neural Networks (ANN)* has been considered as an information processing adaptive system taking advantage of generalisation capabilities [6]. Hence, ANN provided generalised systems with immensely powerful processing capabilities with wide range of solutions for problems in image processing, pattern recognition, and similarly many more. Hence, ANN is a much talked about topic in this paper as it also provided a powerful near-correct set of results in our objective of emotion recognition using facial imagery.

Another technique to solve the problem was looked upon that is the *ORB (Oriented Fast and Rotated BRIEF)* [7], which basically is a fast robust feature detector based on *BRIEF (Binary Robust Independent Elementary Features)* [8], as a facial feature extractor with the *Support Vector Machine (SVM)* as the classifier.

On the third front, the problematic issue of traffic fatalities due to driver drowsiness or drowsy driving was seen which led to the need for systems that promote interaction between driver and vehicle for constant monitoring and supporting resulting into intertwining both the driver and vehicle to conclude into an independent unit while having inter-dependence on each other. This sectorial issue has led to the development of drowsiness detectors using visual perceptive imagery of the vehicle driver's face as data sets to perceive characteristics of

978-1-5386-2361-9/17/$31.00

drowsiness in the driver to interlink the automated emotion recognition systems with that of vehicle safety systems to ensure road safety at a greater degree.

II. Literature Review

A. Affective Recognition using facial imagery

Most of the work in computer based recognition of the human face was completed by detecting and tracking facial features such as nose, eyes, mouth, and figurative outline of the human head [9], attempts at analysing the affective nature or emotion of the human face and drawing effective conclusions about the emotion on display through purely mechanical systems which were intelligent in nature had been conducted since a long time. The development of the FACS [5], is pure testimonial to this fact. As machines are considered the cold component between the parties involved in human computer interaction, hence there had always been the "emotional factor" missing in the interaction between human beings and computational machinery [10] and hence the ever increasing need for developing systems that correlates to human emotions in an effective and full proof manner.

The common denominator in the long term aim to provide seamlessly effective systems is to use a *facial feature extraction techinque* and bind it with *facial feature classification* method culminating in a declarative result of the development of a sufficiently independent system fully capable of carrying of accurate affective recognition of the human face [10].

The recent advancements in human affective recognition from facial images include the proposal of a geometric descriptor for two dynamic approaches where recognition is performed by classifiers such as the Conditional Random Field (CRF) and k-Nearest Neighbours [11], real-time recognition using Raspberry Pi II [1] and the prediction of continuous probability distribution of image emotions which are represented valence-arousal space [12].

B. Use of Neural Networks in facial emotion recognition

Facial feature extraction involved the recognition of several physical facial features such as wrinkles, eyebrows, eyes, skin tone amongst many others. All these facial data led to the formation of data sets. Neural networks were, in such a case ideal to decode the mapping between the data sets involving the facial features and the impression it created. Emotion recogntion using neural networks was first proposed in [13] where the average recognition rate was 85 percent on the 90 images used as the primary data set in the experiment. In the work of [14], ANN was considered with the combination of *Hidden Markov Model (HMM)*, where facial emotion classification is performed for the upper and lower parts of the static image, separately, yet effectively.

C. Using ORB for expression appearance detection

ORB is a feature detector based on BRIEF [8]. These descriptors are based on calculations made from facial key points, for each of which, there is a feature vector present. The facial descriptor is nothing but the collection of the concatenated generated vectors obtained from each key point. The BRIEF descriptor [8] comprises of an image description generated from a set of points of interest (facial data point). While ORB learns optimal sampling pairs, BRIEF uses randomly chosen sampling pairs. Hence, BRIEF does not follow a specific sampling pattern in the facial areas around the key points, thus sampling pairs were chosen randomly at any point on a patch. BRIEF also differs from ORB in the fact that it does not follow any orientation compensation mechanism it is seen that the idea of ORB is to "steer" the BRIEF in accordance to the key points [15]. The algorithm that was used in the SVM with radial basis function kernel where the classifier input for each image a set of feature descriptors extracted from all the landmarks was provided by the database [15].

D. Drowsiness detection using facial feature data sets

1) Capturing the driver's facial features: The use of ORB in capturing the facial features from imagery available was to generate key points and construct data sets, which was useful for feeding into the SVM for emotion classification.

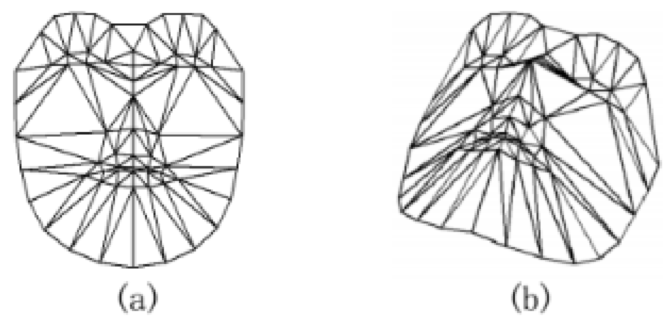

Fig. 1. (a) Specific 2-D model (b) generic 3-D model for AAM

The development of detection and measurement of whether the facial features indicated drowsiness, using image processing without any sensor contact, from a driver's facial image which was video captured is given in [16]. The facial feature extraction in this case was done using Active Appearance Model (AAM) depicted in figure 1 [17], which detected coordinates in 3-dimensional form of measurement points on each frame of the driver's face. Figure 1 is present in [18].

2) Definition of the features of drowsy expression: The inference of drowsiness was drawn by comparing the face muscles of an awake human being to the facial expression of a drowsy individual [19]. An experiment was conducted in [18] where the 9 left facial muscles was measured and the reference states of drowsiness were divided into "Sleeping", "Very Sleepy", "Rather Sleepy", "Sleepy", "Slightly Sleepy" and "Not Sleepy" by adding the "Sleeping" level to Kitajimas trained observer rating scale [20].

III. Experimental Models

A. Implementation of Facial Emotion Recognition Model using ANN

This model takes up two major segments: Facial Image Treatment and ANN propagation algorithm to recognise facial expression [10]. The image treatment primarily consists of a series of phases with the image as an input and morphs it into a new(merged) image to be fed onto the ANN.

Constant analysis of the human face was conducted to produce active and spontaneous categorising of emotional expressions using input of the facial regions around the eyes and mouth and hence vital information was extracted for the construction of the merged image, which basically comprises the features of the face which were geometric (shape), appearance (skin tone and texture, wrinkles, furrows) or points (corners of mouth, eyes), merged into a single new image, which was later resized into using the Nearest Neighbour Interpolation method [10]. The training set for the ANN contained a new matrix which would have a new value replacing the existing values in the minimised image taken from the adjoining areas of the samples and it was important to apply binarisation over the images to fulfil our said objective by applying a pixel threshold related formula given below:

$$V_{n_{pix}} = \left\{ \begin{array}{ll} 0, & if\ V_{c_{pix}} < 230 \\ \left| \frac{V_{c_{pix}} - 255}{255} \right|, & otherwise \end{array} \right\} \tag{1}$$

The value 230 in (1) had been obtained through trial and error depending upon luminous intensity over a set of images.

Once the formation of the merged image was complete, the image was used as input data for ANN which comprised of a back-propagation algorithm with feed-forward architecture.

(1) was used in the algorithm as an input layer in which variables were represented by pixels from the image in the form of 40×30 size in (2).

$$z = \sum_{i=1}^{p} w_i x_i + b = \begin{bmatrix} W^T b \end{bmatrix} \begin{bmatrix} X \\ 1 \end{bmatrix} \tag{2}$$

The input layer of the algorithm made using (1) contains 1200 neurons. In (2), x formed the neuron's activation whereas x_i (pixel) is multiplied by a weight w_i and added after, with b being neuron bias in the initialisation process. The neuron bias b is equal to adding input with weight b and the value 1, (2) reduces to.

$$z = \sum_{i=1}^{p} w_i x_i = W^T X \tag{3}$$

After obtaining the value z which is the the neuron's activation value, was fed into the activation function which return a value between $[0, 1]$ and is a sigmoidal function which is denoted by $\sigma(z)$ in (4).

$$\sigma(z) = \left\{ \begin{array}{ll} 0, & for\ z < 0 \\ 1, & for\ z \geq 0 \end{array} \right\} \quad and \quad \sigma(z) = \frac{1}{1 + e^{-sz}} \tag{4}$$

To capture the nonlinearity of the data, one hidden layer was present to help with one or more neurons while the number of neurons depended upon the number of training samples and was capable of approximating continuous functions. To fix the number of neurons in each hidden layer so that sufficient number of hidden neurons are present, various approaches over the years were used, but for this experiment, approach given in [21] was used. The optimality of the number of hidden units and layers depended upon the complexity of the architecture of network, the total input and output units present, training samples present, training algorithms and to what extent the degree of noise was present in the sample data sets and (3) the output layer contains nodes representing the output variables, whereas the number of emotions denoted the number of neurons that had to be present in the output layer. The training data sets of disgust, happiness and surprise [10] were used. For the creation of the hidden layer present in the structure, with input X_i with neurons present between layers fully inter-connected with weights V_{ij} and W_{jk} and the activation of hidden neurons j is found using

$$z_j = W_j^{h^T} X \tag{5}$$

X, W^h and Z get easily defined in this process.

$$X = \begin{bmatrix} X_1^T \\ X_2^T \\ . \\ . \\ . \\ X_N^T \end{bmatrix}, \ W^h = \begin{bmatrix} W_1^{h^T} \\ W_2^{h^T} \\ . \\ . \\ . \\ W_p^{h^T} \end{bmatrix} \ and$$

$$Z = \begin{bmatrix} z_{11} & z_{12} & \cdots & z_{1_p} \\ z_{21} & z_{22} & \cdots & z_{2_p} \\ . & . & \cdots & . \\ . & . & \cdots & . \\ . & . & \cdots & . \\ z_{N1} & z_{N2} & \cdots & z_{N_p} \end{bmatrix} \tag{6}$$

The elements of matrix Z denoted the activation of hidden neuron j to input sample x_i and the hidden layer output V, system output Y and hidden layer activation Z was found.

$$V = \sigma(z), \quad Z = XW^h \quad and \quad Y = \sigma(VW^0) \tag{7}$$

After this, the ANN got trained by back propagation which involved three steps, feed-forward of input training pattern, the calculation and back-propagation associated error, and weight adjustment. The data obtained was fed forward from the input layer, through the hidden layer to the output and this was to happen without any feedback. As the ANN was initialised, that too randomly, the resulting y was found out by absorbing the input x, the output desired denoted by d was useful in the calculation of the back propagation of associated error $e = d - y$, the goal to minimise the cost function thoroughly depicted in [10]. The data was actualised through law for hidden neuron weights calculated in (10).

$$w_{ij}^h(n+1) = w_{ij}^h(n) + \alpha(n) x_i \sigma_j'(z_j) \sum_l e_l w_{jl}^0 \tag{8}$$

Gradient descent was searched by back propagation by the feed forward error back propagation learning algorithm. The portion of error correction had been calculated and the weights for all layers were judged simultaneously, it was necessary to use the N test samples $x_i,\ d_i$ in order to judge the weights. The training data sets that correlated to each facial expression required per individual are present in a merged image.

The three layers that were required in order to train the ANN and the "Sigmoid" for transfer function is important as the data set is normalised, along with "Bias Neurons" option. The weights when computed made the ANN ready for selecting the emotion present in the merged image. The flowchart for the entire Facial Emotion Detection using ANN is given figure 2 [10].

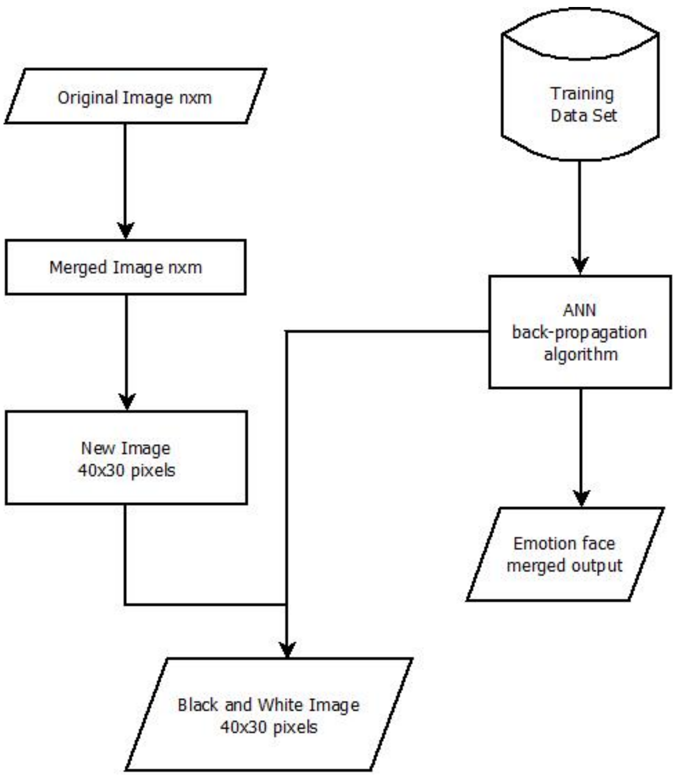

Fig. 2. The flowchart for the entire experiment of Facial Emotion Detection using ANN

B. Using static approach with ORB + SVM

To gauge the accuracy of the face descriptor, ORB, for its use as a feature descriptor for facial emotional expression recognition, experiments were conducted in [15].

A radial basis function kernel attached with SVM was the classification algorithm used to estimate the expression with the implementation done using OpenCV implementation for ORB and the LIBDVM library was used for using the SVM algorithms.

The first experiment included seven different facial emotion categories which were used to find out using the dynamic approach which were anger, disgust, fear, contempt, happiness sadness and surprise while neutral emotion was also considered as an emotion category in the second experiment. Sixty-eight facial landmarks were used as key points, while for each landmark, $n = 256$ binary tests were computed to extract a feature vector.

TABLE I

CONFUSION MATRIX OF FACIAL EXPRESSION RECOGNITION FOR THE ORB FEATURES ON THE SEVEN EMOTIONS

	An	Co	Di	Fe	Ha	Sa	Su
An	**57.8**	2.2	11.1	2.2	0	26.7	0
Co	11.1	**66.7**	0	0	0	22.2	0
Di	1.7	3.4	**78**	3.4	3.4	10.2	0
Fe	8	4	16	**40**	12	8	12
Ha	0	5.8	2.9	1.4	**89.9**	0	0
Sa	0	7.1	14.3	0	0	**78.6**	0
Su	1.2	1.2	9.6	2.4	1.2	6	**78.3**

TABLE II

CONFUSION MATRIX OF FACIAL EXPRESSION RECOGNITION FOR THE ORB FEATURES CONSIDERING THE NEUTRAL EXPRESSION

	An	Co	Di	Fe	Ha	Sa	Su	Ne
An	**80**	2.2	3.6	2.2	0	2.2	0	8.9
Co	5.6	**61.1**	0	0	5.6	5.6	0	22.2
Di	3.4	0	**91.5**	1.7	1.7	0	0	1.7
Fe	4	0	8	**60**	12	0	8	8
Ha	0	2.9	1.4	1.4	**94.2**	0	0	0
Sa	3.6	0	3.6	0	0	**71.4**	0	21.4
Su	0	1.2	0	2.4	0	1.2	**95.2**	0

The last frame of each sequence was the only portion used in the basic training expression. Each frame of the sequences, for the tests, was classified and class of the sequence was determined according to the most numerous expression that was present. Confusion matrix results are shown in Table I.

The first frame of a sequence from each subject in the database was added to the training set when the neutral expression was added to the set of basic facial expressions. Each sequence in the database corresponds to one of the seven basic facial expressions except neutral, the frames that were omitted which were classified as neutral as most numerous expression in the sequence was determined. As and when neutral classification was done for all the image frames, the entire sequence was classified as neutral and not omitted. Table II denoted confusion matrix with percentages of sequence considered as neutral and it was seen that the percentage of sequences whose frames were all misclassified as neutral whose actual class was indicated by the corresponding row by the last column.

The rate of the overall performance, which was calculated by finding out the average of the confusion matrix diagonal percentage values, was 69.9% at first, and 79.1% when the neutral emotion was added on later.

C. Detection of drowsiness levels

Figure 3 is given in [18] and seventeen points for the procedure to detect drowsiness through facial imagery were adopted, which included ten points of the left and right eyebrows (five points on each eyebrow), four points of the left and right eyelids (two on each eyelid), two points on the left and right edges of the mouth (1 on each side) and one point on the lower lip. Individual differences were computed in the reference of

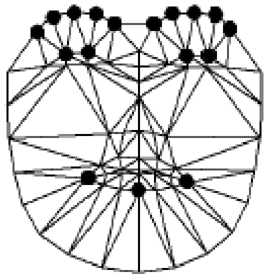

Fig. 3. Seventeen points for measurement for drowsy expression detection

the state of awake expression which was reduced on defining the position of the points in the awake state as reference points. The use of k-Nearest Neighbour method was used, for the detection of drowsiness facial expression and was one the of the pattern classification methods. The technique consisted of the six level drowsy expression features by using the prebuilt database. The driver's facial imagery was compared to the entire database, and the similarities of each comparison made were used the detect the level of drowsiness present. This method was able to detect expressions of drowsiness after every 5 second time interval window, which was much less than the 30 second time interval window used by the similarity based method, which was able to detect signs of drowsiness with a higher time resolution by using trends in the change of facial expression done in the works of [22]. The every 5 second time interval included 150 frames with feature data used in this method. The 5 second block time interval was the bare minimum sampling time for a observer rating which had been trained for facial expressions [20].

The main crux line of the averaging leads to the formation of a conclusion that successfully distinguished the difference between "eye closure due to continuous blinking", which was a normal awake state expression characteristic and "eye closure based on drowsiness", which was difficult to distinguish from a single frame.

The facial imagery of all participants was used for drowsiness detection which was performed using leave-one-out cross validation which was conducted offline and was done on the facial images of all 13 participants, all of whom, did fall asleep at some point of time. In leave-out-one procedure, the method included using data of any one of 12 out of the 13 participants selected arbitrarily as training data and this was done repeatedly. The Root Mean Square Errors (RMSE) of all the 13 participants demonstrate that the technique detects drowsiness with RMSE with an average of 0.91 [23] as given in Table III.

TABLE III
ROOT MEAN SQUARE ERRORS OF DROWSINESS DETECTION

Participant #	Root Mean Square Error (RMSE)
1	1.06
2	0.90
3	1.14
4	0.91
5	0.78
6	1.11
7	0.71
8	1.00
9	0.82
10	0.81
11	0.93
12	0.77
13	0.85
Average	0.91
SD	0.14

IV. Conclusion and Future Scope

The experiment results showcase the facial emotion detection system using the complete merged of the images present in the database and using it as an input to the ANN, the results pretty convincing.

Table IV denoted the confusion matrix of the emotion recognition system based the facial imaging inputs. The overall performance of the classifier was 83.3% while the diagonal components reveal that all emotions were revealed with more than 75% accuracy.

TABLE IV
CONFUSION MATRIX FOR FACIAL EMOTION RECOGNITION SYSTEM USING COMPLETE MERGED IMAGE

	Anger	Disgust	Fear	Happy	Sad	Surprise
Anger	75.00	0.00	0.00	16.67	0.00	0.00
Disgust	0.00	83.33	8.33	0.00	16.67	0.00
Fear	0.00	0.00	75.00	0.00	0.00	0.00
Happy	0.00	0.00	0.00	83.33	0.00	0.00
Sad	25.00	16.67	0.00	0.00	83.33	0.00
Surprise	0.00	0.00	16.67	0.00	0.00	100.00

Hence these results showcases the great potential and the fact that artificial neural networks to detect and recognise real life human emotion and expressions. Also, in [10] ANN was used to detect emotions using data consisting of the facial regions of the eyes and mouth only, thus a reduced amount of data sets required as other parts of the face containing unnecessary information which is removed readily. Future scope in this field includes the development of dynamic neural networks, which was the state of the network evolving from time to time from the initial state to the final state corresponding with the face containing the desired emotion. With the dynamic neural network, the information provided to manage relevant and important information and construct a model that supported the an autonomous interviewer with a fully independent decision making system, which interacts with the users on its own. The emotion showcased by the human faces would be the main concern, which was analysed and used as a variable in the decision making model [10].

Moving onto emotion classification using a static approach by chalking out ORB as a facial descriptor by generating information to be fed into the SVM and this method produces an accuracy of 79.1% after the inclusion of the neutral expression. The ORB descriptor had a clear advantage over the dynamic approach modelled by the Conditional Random Fields (CRF) [15], as the ORB is applied to a single frame and does not depend upon the entire motion performed to reach the expression. There are several approach to merging the static and dynamic approach to construct a model for on-the-go recognition with possibly better accuracy.

This research article projects the recent methods of emotion recognition including that of the drowsy expression. Drowsiness detection is a critical real life problem to help reduce road fatalities and hence a drowsiness detection model that traces 17 facial points on the human face by image processing and classifies by the rating the driver's drowsiness on a 6 level scale and the result successfully corresponds to a RMSE of less than 1.0 level [18]. The future of this work lies in continuing the experiment in a motion based driving simulation environment which will determine the practical viability of the study conducted. In that phase, the challenges will be more as the parameters to be considered would increase considerably and the most complex one would be to differ drowsiness emotion to various complexly mixed human expressions. This work also necessitates the development of a feedback system to succeed in the construction of an intelligent safety vehicle system that relieves drivers struggling against drowsiness, hence reducing the gap that lies in the communication between the human and vehicle.

The problem of development of automated real time emotion recognition systems promises to play a huge role in entertainment, combating autism, promoting digital learning, reducing traffic fatalities and development of intelligent monitoring security systems. Here, in this paper, it is shown that machine learning and deep learning models can be increasingly used to improve the accuracy of the trained models to recognise human behavioural emotions after the generation of accurate data sets through facial image analysis by the collection of data points of the various regions of the human face through data consisting solely of imagery. In the methods showcased here, we can see that machine learning models can be trained by using appropriate data sets in an attempt to fulfill the long term goal of building systems that serve a greater automated role in computational systems that recognise human emotions or moods, for mankind.

References

[1] P. Suja, S. Tripathi *et al.*, "Real-time emotion recognition from facial images using raspberry pi ii," in *Signal Processing and Integrated Networks (SPIN), 2016 3rd International Conference on*. IEEE, 2016, pp. 666–670.

[2] M. Pantic and L. J. M. Rothkrantz, "Automatic analysis of facial expressions: The state of the art," *IEEE Transactions on pattern analysis and machine intelligence*, vol. 22, no. 12, pp. 1424–1445, 2000.

[3] A. Bhattacharya, D. Choudhury, and D. Dey, "Emotion recognition from facial image analysis using composite similarity measure aided bidimensional empirical mode decomposition," in *Control, Measurement and Instrumentation (CMI), 2016 IEEE First International Conference on*. IEEE, 2016, pp. 336–340.

[4] Z. Zeng, M. Pantic, G. I. Roisman, and T. S. Huang, "A survey of affect recognition methods: Audio, visual, and spontaneous expressions," *IEEE transactions on pattern analysis and machine intelligence*, vol. 31, no. 1, pp. 39–58, 2009.

[5] P. Ekman, W. Friesen, and J. Hager, "Facial action coding system (facs): Manual. 2002," *A Human Face, Salt Lake City (USA)*.

[6] K. Madani, "Artificial neural networks based image processing & pattern recognition: From concepts to real-world applications," in *Image Processing Theory, Tools and Applications, 2008. IPTA 2008. First Workshops on*. IEEE, 2008, pp. 1–9.

[7] E. Rublee, V. Rabaud, K. Konolige, and G. Bradski, "Orb: An efficient alternative to sift or surf," in *Computer Vision (ICCV), 2011 IEEE international conference on*. IEEE, 2011, pp. 2564–2571.

[8] M. Calonder, V. Lepetit, C. Strecha, and P. Fua, "Brief: Binary robust independent elementary features," *Computer Vision–ECCV 2010*, pp. 778–792, 2010.

[9] M. A. Turk and A. P. Pentland, "Face recognition using eigenfaces," in *Computer Vision and Pattern Recognition, 1991. Proceedings CVPR'91., IEEE Computer Society Conference on*. IEEE, 1991, pp. 586–591.

[10] J. G. Rázuri, D. Sundgren, R. Rahmani, and A. M. Cardenas, "Automatic emotion recognition through facial expression analysis in merged images based on an artificial neural network," in *Artificial Intelligence (MICAI), 2013 12th Mexican International Conference on*. IEEE, 2013, pp. 85–96.

[11] D. Acevedo, P. Negri, M. E. Buemi, F. G. Fernández, and M. Mejail, "A simple geometric-based descriptor for facial expression recognition," in *Automatic Face & Gesture Recognition (FG 2017), 2017 12th IEEE International Conference on*. IEEE, 2017, pp. 802–808.

[12] S. Zhao, H. Yao, Y. Gao, R. Ji, and G. Ding, "Continuous probability distribution prediction of image emotions via multitask shared sparse regression," *IEEE Transactions on Multimedia*, vol. 19, no. 3, pp. 632–645, 2017.

[13] H. Kobayashi and F. Hara, "Facial interaction between animated 3d face robot and human beings," in *Systems, Man, and Cybernetics, 1997. Computational Cybernetics and Simulation., 1997 IEEE International Conference on*, vol. 4. IEEE, 1997, pp. 3732–3737.

[14] T. Hu, L. C. De Silva, and K. Sengupta, "A hybrid approach of nn and hmm for facial emotion classification," *Pattern Recognition Letters*, vol. 23, no. 11, pp. 1303–1310, 2002.

[15] D. Acevedo, P. Negri, M. E. Buemi, and M. Mejail, "Facial expression recognition based on static and dynamic approaches," in *Pattern Recognition (ICPR), 2016 23rd International Conference on*. IEEE, 2016, pp. 4124–4129.

[16] T. Kimura, K. Ishida, and N. Ozaki, "Feasibility study of sleepiness detection using expression features," *Review of Automotive Engineering*, vol. 29, no. 4, pp. 567–574, 2008.

[17] T. F. Cootes, G. J. Edwards, and C. J. Taylor, "Active appearance models," *IEEE Transactions on pattern analysis and machine intelligence*, vol. 23, no. 6, pp. 681–685, 2001.

[18] S. Hachisuka, "Human and vehicle-driver drowsiness detection by facial expression," in *Biometrics and Kansei Engineering (ICBAKE), 2013 International Conference on*. IEEE, 2013, pp. 320–326.

[19] K. Ishida, A. Ichimura, and M. Kamijo, "A study of facial muscular activities in drowsy expression," *Kansei Engineering International Journal*, vol. 9, no. 2, pp. 57–66, 2010.

[20] H. Kitajima, N. Numata, K. Yamamoto, and Y. Goi, "Prediction of automobile driver sleepiness(1 st report, rating of sleepiness based on facial expression and examination of effective predictor indexes of sleepiness)," *Nippon Kikai Gakkai Ronbunshu, C Hen/Transactions of the Japan Society of Mechanical Engineers, Part C*, vol. 63, no. 613, pp. 3059–3066, 1997.

[21] J. Ke and X. Liu, "Empirical analysis of optimal hidden neurons in neural network modeling for stock prediction," in *Computational Intelligence and Industrial Application, 2008. PACIIA'08. Pacific-Asia Workshop on*, vol. 2. IEEE, 2008, pp. 828–832.

[22] E. Vural, M. Çetin, A. Erçil, G. Littlewort, M. Bartlett, and J. Movellan, "Automated drowsiness detection for improved driving safety," 2008.

[23] S. Hachisuka, K. Ishida, T. Enya, and M. Kamijo, "Facial expression measurement for detecting driver drowsiness," *Engineering Psychology and Cognitive Ergonomics*, pp. 135–144, 2011.