Facial Expression Recognition using Facial Landmark Detection and Feature Extraction on Neural Networks

Fuzail Khan
Department of Electronics
and Communication Engineering
National Institute of Technology Karnataka, Surathkal
Mangalore, India 575025

***Abstract*—The proposed framework in this paper has the primary objective of classifying the facial expression shown by a person using facial landmark detection and feature extraction. These classifiable expressions can be any one of the six universal emotions along with the neutral emotion. After initial facial detection, facial landmark detection and feature extraction are performed (where in the landmarks were determined to be the fiducial features: the eyebrows, eyes, nose and lips). This is primarily done using the Sobel horizontal edge detection method and the Shi Tomasi corner point detector. This leads to input feature vectors being formulated and trained into a Multi-Layer Perceptron (MLP) neural network in order to classify the expression being displayed. Facial Expression Recognition (FER) is a significant step in reaching the eventual goal of artificial intelligence. If efficient methods can be brought about to automatically recognize these expressions, major advances may be achieved in computer vision.**

***Index Terms*—Facial expression recognition, Facial landmark detection, Facial feature extraction, neural networks.**

I. Introduction

The implementation can be broadly categorized into four stages: face location determination stage, facial landmark detection stage, feature extraction stage and emotion classification stage. An appropriate facial database was to be obtained which serves as our training and our testing data set, essentially consisting of humans displaying labelled emotions in the images [1]. The first stage deals with face detection algorithms that are to be implemented for the facial recognition, which mainly deals with image pre-processing and normalizing the images to eliminate redundant areas. In the second stage, it is stated that our facial fiducial landmarks: eyebrows, eyes, nose and mouth, are identified as the critical features for emotion detection and their feature points are hence extracted to recognize the corresponding emotion. These feature points are extracted from the selected feature regions with the use of an edge detection and a corner point detection algorithm, which are experimented with an ensemble of detection methods. The Sobel horizontal edge detection method and the Shi Tomasi corner point detection were eventually used for the purpose. The input feature vectors were then calculated out of the facial feature extracted points obtained. In the third stage, these input feature vectors are given as input to the MLP neural network that is trained to then classify what emotion is being shown by the human.

II. Related work

Facial Expression Recognition (FER) systems have been implemented in a multitude of ways and approaches. The majority of these approaches have been based on facial features analysis while the others are based on linguistic, paralanguage and hybrid methods. Ghimire et al. [2] used the concept of position-based geometric features and angle of 52 facial landmark points. First, the angle and Euclidean distance between each pair of landmarks within a frame are calculated, and then successive subtraction between the same in the next frame of the video, using a SVM on the boosted feature vectors. The appearance features are usually extracted from the global face region [3] or different face regions containing different types of information [4,5]. Happy et al. [3] utilized features of salient facial patches to detect facial expression. This was done after extracting facial landmark features and then using a PCA-LDA hybrid approach for dimensionality reduction and performance improvement. They used a radial basis function kernel for SVM multi-class classification.

For hybrid methods, some approaches [6] have combined geometric and appearance features to complement the positive outcomes of each other and in fact, achieve better results in certain cases. In video sequences, many systems [2,7,8] are used to measure the geometrical displacement of facial landmarks between the current frame and previous frame as temporal features, and extracts appearance features for the spatial features. Szwoch et al. [9] recognized facial expression and emotion based only on depth channel from the Microsoft Kinect sensor without using a camera. Local movements in the facial region are seen as features and facial expressions are determined using relations between particular emotions. Similarly, Sujono et al. [10] used the Kinect sensor to detect the face region and for face tracking based on the Active Appearance Model (AAM). Polikovsky et al. [11] presented facial micro-expression recognition in videos captured from 200 frames per second (fps) high speed camera. This method divides the face regions into certain localized regions, and then a 3D Gradients Orientation Histogram is generated from the motion in each local region.

Shen et al. [12] used infra-thermal videos by extracting horizontal and vertical temperature differences from different face sub regions. The Adaboost algorithm with the weak classifiers of k-Nearest Neighbor is used for Expression Recognition. Some researchers [9,10,12,13,14] have tried to recognize facial emotions using infrared images instead of images illuminated by visible light because the degree of dependence of visible light images on illumination is considerably higher. Zhao et al. [13] used near-infrared (NIR) video sequences and LBP-TOP (Local Binary Patterns-Three Orthogonal Planes) feature descriptors. This study uses component-based facial features to combine geometric and appearance information of face. An SVM and sparse representation classifiers are used for the emotion classification. Conventional FER systems in general use considerably lower processing power and memory when balanced with deep learning based approaches and are thus, still being researched for use in real time mobile systems because of their low computational power and high degree of reliability and precision.

III. Data Collection

After a comprehensive search of databases that suited our purpose, the Karolinska Directed Emotional Faces (KDEF) dataset [1] was used to test our approach. It was developed at Karolinska Institutet, Department of Clinical Neuroscience, Section of Psychology, Stockholm, Sweden. It consists of a set of 4900 images of 70 subjects - 35 male and 35 female, each showing the 6 basic expressions and 1 neutral expression. Each expression being photographed twice from 5 different angles. The original image size is 562 x 762 pixels. A sample of images showing all 7 expressions is shown in Figure 1.

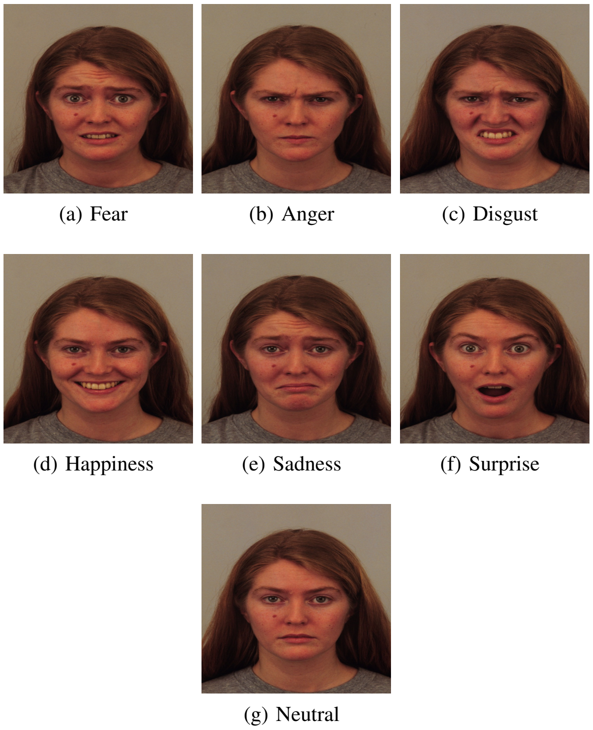
The data is partitioned as 90:10 for training data:testing data. This ratio was chosen as the objective is to train the neural network with maximum data along with the fact that 490 images are relatively enough to test the accuracy of the algorithm.

IV. Image pre-processing

On getting the input image, the first step is to perform image pre-processing for the purpose of removing unwanted noise from the image and for enhancing the contrast of the image. A Low Pass 3x3 Gaussian filter was applied which helped smoothen the image and normalize gradient intensity values. Contrast Adaptive Histogram equalization was then carried out for illumination corrections. Normally, pre-processing would be to ensure uniform shape and size of the input images. This would not really apply to the KDEF database images as the facial orientation and location of the faces in each picture is uniform and hence, removes the need for any database specific pre-processing.

V. Facial Detection

In the facial detection stage, the objective is to find the face so as to limit our Region of Interest (RoI) such that all further processing takes place from that RoI onwards. This step of facial detection was accomplished using the now very reliable method of Haar classifiers. Haar feature-based cascade classifiers is an effective object detection method proposed by Paul Viola and Michael Jones [15]. It is a machine learning based approach where a cascade function is trained from a lot of positive and negative images.



(a) Fear (b) Anger (c) Disgust

(d) Happiness (e) Sadness (f) Surprise

(g) Neutral

Fig. 1: Sample of images from the KFED facial database with 6 basic emotions and 1 neutral emotion being displayed.

Each feature is a single value obtained by subtracting sum of pixels under white rectangle from sum of pixels under black rectangle. It determines features that best classifies the face and non-face images. Final classifier is a weighted sum of these weak classifiers as Adaboost combines many weak classifiers into one single strong classifier.

The features are then applied in several stages leading to a cascade of classifiers. The cascaded stages are implemented with 1, 10, 25, 25 and 50 features in the first five stages.

After the face is detected by this approach, the image is resized to only the face region and facial landmark detection takes place from this region onwards as seen in Fig. 4 b).

VI. Landmark detection

A. Eyes detection

The purpose is to get an RoI of the right eye and the left eye that is then to be used to extract feature points corresponding to the eyes. The most accurate alternatives were:

a) Using Haar classifiers - where both the eyes will be detected separately using Haar classifiers trained for each eye.

b) Using Circular Hough Transform - This will detect circles,

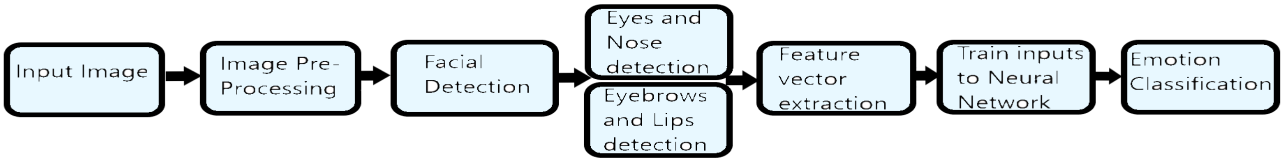
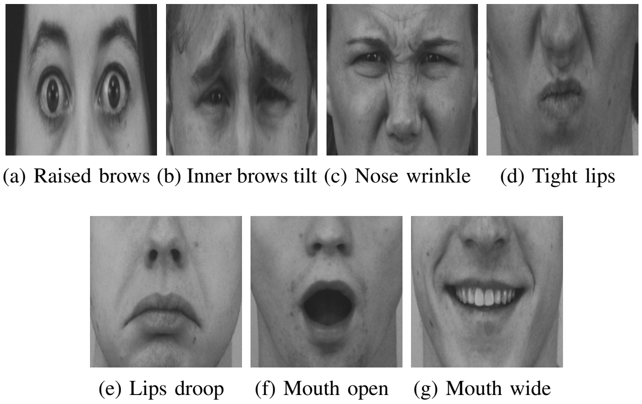


Fig. 2: Flowchart of the proposed FER methodology


(a) Raised brows (b) Inner brows tilt (c) Nose wrinkle (d) Tight lips

(e) Lips droop (f) Mouth open (g) Mouth wide

Fig. 3: Sample of images from the KFED facial database showing marked visible changes in facial landmarks for varied expressions.

and hence will detect the pupil of the eyes.
The characteristic equation of a circle of radius r and center (a,b) is given by:

$$(x-a)^2 + (y-b)^2 = r^2 \tag{1}$$

This circle can be described by the two following equations:

$$x = a + rcos\theta \tag{2}$$

$$y = b + rsin\theta \tag{3}$$

Thus, the role of the Hough transform is to search for the triplet of parameters (a, b, r) which determines the points.
Of the two alternatives, it was found that the Circular Hough Transform was a better method as through this, we manage to localize the eye center or the pupil. We then build a bounding box around the eyes taking the Hough Circle as the centroid of the box, using image gradient derivatives due to the sharp contrast between the eye white region and the pupil. Also, the Circular Hough Transform gave us the flexibility to decide parameters such as max. radius and min. radius of the circle to be detected giving us more control and so, this method was chosen, as seen in Fig. 4 c).

B. Nose detection

Using the knowledge that the nose is right below the eyes, we use the evidence of the already detected eyes to start resizing and tracking vertically downwards. The simplest method that gave us a 97.9% detection rate was using the trained Haar classifier for the nose (Fig. 4 c)) Other methods tried were using Histogram of Oriented Gradients (HOG) and SURF features, two very common methods used in object recognition and classification.

C. Eyebrows and lips detection

Using the knowledge that the eyebrows are above the eyes and the lips are below the nose, and since we already have the location of the eyes and the nose, we process only on those corresponding RoI's to ease computation. The eyebrows and upper lip always produce a distinct edge which can be detected using a horizontal edge detector. Eyebrow detection would be accurately performed by these edge detection algorithms. Common edge detection algorithms include Sobel, Canny, Prewitt and Roberts.
The Sobel-Feldman operator is based on convolving the image with an integer-valued filter in the horizontal and vertical direction that give us two images which at each point contain the vertical and horizontal derivative approximations of the source image. For the purpose of detecting eyebrows and the upper lip, the Sobel horizontal edge operator was decided upon.
After performing the edge detection, surplus edges were present. A trial-and-error process had to be performed to check the effectiveness of each thresholding method. The basis of this effectiveness was how well the excess edges were refined out leading us to focus on the relevant feature points. The Otsu's thresholding turned out to be the best approach in this sense, trumping other forms of binary thresholding techniques. It's bimodal histogram nature facilitates the purpose. Further, these false components having an area less than a specified threshold were removed. Finally, maximising the objective area function, the connected component with the maximum area which was just below the nose region was selected as upper lip region and just above the eyes were selected as the eyebrows.

VII. Feature corner point extraction

From the detection windows obtained for the eyes and nose, and then the eyebrows and lips, we use these RoI windows to obtain the corner point features, which will be used to fed as feature input vectors to the neural network. For this purpose, the Shi Tomasi corner point detector is implemented.
The Shi Tomasi method determines which windows produce very large variations in intensity when moved in both X and Y directions, thus computing the X and Y gradients. With each such window found, a score R is computed.
With each corresponding window, we have a window function = w(x,y).
We calculate M = w(x,y) x I where I =

$$\begin{bmatrix} \Sigma_{(x,y)} {I_x}^2 & \Sigma_{(x,y)} I_x I_y \\ \Sigma_{(x,y)} I_y I_x & \Sigma_{(x,y)} {I_y}^2 \end{bmatrix}$$

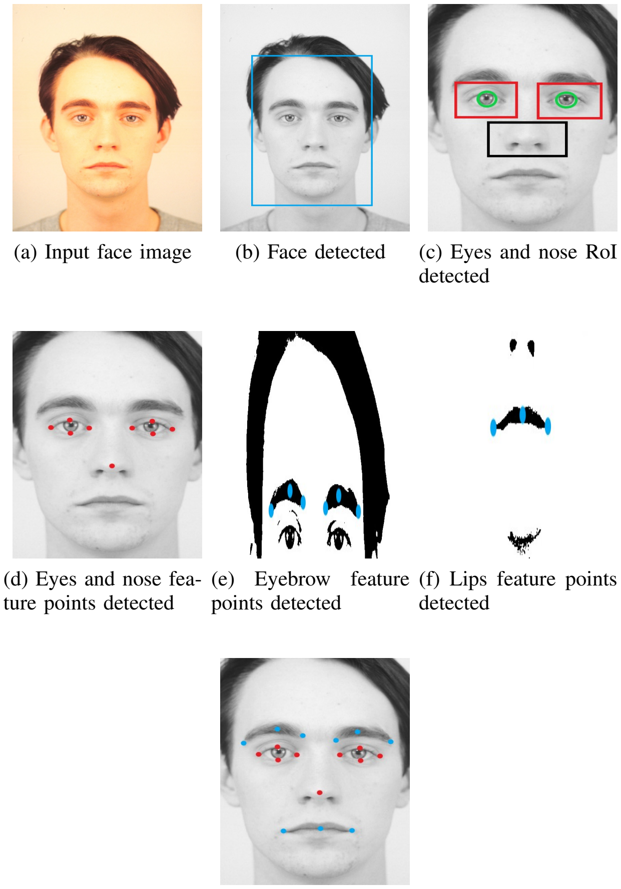
where I_x and I_y are image derivatives in the x and y direction respectively.
The value R for each window:

$$R = min(\lambda_1, \lambda_2) \tag{4}$$

where λ_1 and λ_2 are the eigen values of M.
After applying a threshold to R, important corners are selected and marked. Tuning parameters such as the minimum quality of image corners, minimum Euclidean distance between the corners and the window size for computing the derivative covariation matrix, the corner points for the fiducial landmarks are extracted.
The selected feature corner points are :
For the eyes : 8 feature points - Left eye upper corner (F1), Left eye lower corner (F2), Left eye left corner (F3), Left eye right corner (F4), Right eye upper corner (F5), Right eye lower corner (F6), Right eye left corner (F7), Right eye right corner (F8).
For the nose : 1 feature point (F9).
For the eyebrows : 6 feature points - Left eyebrow left corner (F10), Left eyebrow right corner (F11), Left eyebrow point directly above left eye centre (F12), Right eyebrow left corner (F13), Right eyebrow right corner (F14), Right eyebrow point directly above right eye centre (F15).
For the lips : 3 feature points - Left corner (F16), right corner (F17) and point directly below nose centre (F18).
The final image is as shown in Fig 4.

VIII. Finding feature vectors

Now that we have the feature points extracted from the input image (F1-F18), the next step is to decide the input feature vectors that need to be fed into and used to train our neural network. On the basis of these input feature vectors, the neural Network will hence learn from these and use these inputs to make classification decisions of the final output emotion. The chosen feature vector inputs are as seen in Table 1.



(a) Input face image

(b) Face detected

(c) Eyes and nose RoI detected

(d) Eyes and nose feature points detected

(e) Eyebrow feature points detected

(f) Lips feature points detected

(g) Final image with all feature points shown

Fig. 4: Various stages of facial landmark detection and feature extraction being shown.

TABLE I: Input feature vectors

Definition	Formula
Left eye Height	V1 = F1-F2
Left eye Width	V2 = F4 - F3
Right eye Height	V3 = F5 - F6
Right eye Width	V4 = F8- F7
Left eyebrow width	V5 = F11 - F10
Right eyebrow width	V6 = F14 - F13
Lip width	V7 = F17 - F16
Left eye upper corner and left eyebrow center dist.	V8 = F12 - F1
Right eye upper corner and right eyebrow center dist.	V9 = F15 - F5
Nose centre and lips centre dist.	V10 = F9 - F18
Left eye lower corner and lips left corner dist.	V11 = F2 - F16
Right eye lower corner and lips right corner dist.	V12 = F6 - F17

IX. Training the neural Network

A. *Neural networks and back propagation*

Neural networks have proved their ability in the recent past to deliver simple and powerful solutions in areas relating to signal processing, artificial intelligence and computer vision. A neural network is represented by weighted interconnections between layers of nodes or neurons. The Back-propagation neural network is the most widely used neural network algorithm due to its simplicity, together with its universal approximation capacity.

The back-propagation algorithm defines a systematic way to update the synaptic weights of multi-layer perceptron (MLP) networks. The supervised learning is based on the gradient descent method, minimizing the global error on the output layer. The learning algorithm is performed in two stages: feed-forward and feed- backward. In the first phase the inputs are propagated through the layers of processing elements, generating an output pattern in response to the input pattern presented. In the second phase, the errors calculated in the output layer are then back propagated to the hidden layers where the synaptic weights are updated to reduce the error. This learning process is repeated until the output error value, for all patterns in the training set, are below a specified value.

B. The neural network configuration

The MLP neural network consists of 4 layers - an input layer, 2 hidden layers and an output layer. Since the input feature vector V [V1 V2 ... V12] is 12 dimensional, the first hidden layer has 12 neurons and the second hidden layer has 7, the number of classified emotions.

Regarding the activation functions for the nodes, non-linear activation functions are used for the neurons in an MLP network. To decide the activation function, we need to examine the purpose and the output of the neural network :

P(Y_i/x) = Probability of Emotion (Y_i) given the input image x, where i ϵ [0,6] corresponding to the 7 output emotions.

Keeping this in mind, it would be appropriate for us to use the Sigmoid logistic activation function.

Sigmoid function :

$$\phi(z) = \frac{1}{(1 + e^{-z})} \tag{5}$$

For a multi-class classification, we will use the softmax Sigmoid activation function where a k-dimensional vector is the output for a given k-class classification where each of the k values $\epsilon(0, 1)$ and sum up to 1 :

$$\sigma : R^k \longrightarrow \left\{ \sigma \; \epsilon \; R^k \mid \sigma_i > 0 \; ; \quad \Sigma_{i=1}^{K} = 1 \right\} \tag{6}$$

$$\sigma(z)_j = \frac{e^{z^j}}{\Sigma_{k=i}^{K} e^{z^j}} \; ; \quad j\epsilon[1, k] \tag{7}$$

Thus, the final emotion classified =
max(P(Y_i/x)) $\forall$ i ϵ [0,6] i.e the maximum of the probabilities of all the 7 emotions given the image x.

After the process of parameter tuning, optimization and regularization, the neural network was configured by:
the optimal learning rate = 0.5, the target of error = 0.0001 and maximum epoch = 1000.

X. Results and Conclusion

After training the neural network as explained above, we use our testing set of images to check the performance of the FER system. The results of the test have been presented as a confusion matrix as shown in Table 2 and the false positive detection rates per emotion as shown in Table 3.

TABLE II: Confusion matrix of emotion classification

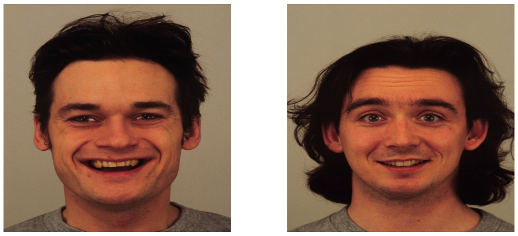
I/O	Happiness	Anger	Disgust	Surprise	Fear	Sadness	Neutral
Happiness	98.2%	0%	0%	0%	1.5%	0%	0.3%
Anger	0%	85.6%	9.3%	2.6%	1.7%	0.8%	0%
Disgust	0%	4.7%	84.9%	1.1%	1.1%	7.0%	1.2%
Surprise	0%	0.4%	1.1%	95.8%	1.3%	1.4%	0%
Fear	1.2%	1.3%	4.2%	1.7%	86.5%	2.4%	2.7%
Sadness	0%	1.1%	0.4%	0.4%	9.6%	86.6%	1.9%
Neutral	0.6%	0.5%	2.1%	0%	2.4%	4.3%	90.1%

From Table 2 and Table 3, we see that the emotions of happiness and surprise are being detected relatively well with 95%+ true positive success rates, which indicates that the facial reactions to when a person is happy and surprised are more uniform than the other emotions leading to high success rates. The emotions of anger, sadness, disgust and fear have their success rates in the 84-86% range and have an overlap among almost all of the other emotions indicating that people may have the same facial reaction for two different emotions as well as having different facial reactions for the same emotions, which leads to the overlap among these classified emotions, while the neutral emotion is being recognized quite well with a 90.1% success rate.

TABLE III: False positive rate per emotion

Emotion	False positive rate
Happiness	1.8%
Anger	14.4%
Disgust	15.1%
Surprise	4.2%
Fear	13.5%
Sadness	13.4%
Neutral	9.9%

It is well known that expression is inherently subjective and different people react in varied manners to different emotions. For example, person A looking afraid might look similar to person B looking disgusted. Our conclusions of happiness and surprise having lesser false positive detection rates can be used to further analyse this concept. As a preliminary act of scratching the surface, we can focus on feature vectors associated with the lips. As noted in Table 3, these relevant feature vectors would be V7, V10, V11 and V12. These feature vectors are essentially the euclidean distance between two extracted feature points. These can be used to differentiate between different subjects showing varied degrees of expression. With the intention of maintaining a 50-50 split, we shall use the median to split the data on these feature vectors. This concept applies with the implicit assumption that, for instance, the wider the smile, hence the greater the feature vector and the more expressive the person. Fig. 5 illustrates our notion of subjective expressions in the data.



(a) A smile to a greater degree

(b) A smile to a lesser degree

Fig. 5: The emotion of happiness being conveyed with different extents of expression.

Experimenting with these two sets of data showed better results among the two groups, especially with the emotions of anger and fear. As was expected, the 'more expressive' people display a higher degree of correlation in expressing emotions among themselves and the same applies to the 'less expressive' group. Further exploration into the individuality and personal nature of facial expressions can better recognize emotions going forward.

On a larger note, a major prospective application would be to maximise use of FER systems and the like in personal assistants like Siri and Alexa. Since these personal assistants will be constantly learning only on their respective user and since that user has certain mannerisms and behaviours unique to himself, the ability to correctly identify his or her moods and emotions will be significantly improved.

Considering future improvements, more data would lead to better results and so, the number of images in the facial expression databases is hence, a limiting factor. Also, the incidence of facial hair, occlusions and the like have not been taken into account and so, further analysis and studies are required to better performance by dealing with these subjects.

References

[1] Lundqvist, D., Flykt, A., hman, A. (1998). The Karolinska Directed Emotional Faces - KDEF, CD ROM from Department of Clinical Neuroscience, Psychology section, Karolinska Institutet, ISBN 91-630-7164-9

[2] Ghimire, D.; Lee, J. Geometric feature-based facial expression recognition in image sequences using multi-class AdaBoost and support vector machines. Sensors 2013, 13, 77147734. [CrossRef] [PubMed]

[3] Happy, S.L.; Routray, A ,Automatic Facial Expression Recognition Using Features of Salient Facial Patches in IEEE Transactions on Affective Computing - May 2015, DOI 10.1109.

[4] Siddiqi, M.H.; Ali, R.; Khan, A.M.; Park, Y.T.; Lee, S. Human facial expression recognition using stepwise linear discriminant analysis and hidden conditional random fields. IEEE Trans. Image Proc. 2015, 24, 13861398. [CrossRef] [PubMed]

[5] Khan, R.A.; Meyer, A.; Konik, H.; Bouakaz, S. Framework for reliable, real-time facial expression recognition for low resolution images. Pattern Recognit. Lett. 2013, 34, 11591168. [CrossRef]

[6] Ghimire, D.; Jeong, S.; Lee, J.; Park, S.H. Facial expression recognition based on local region specific features and support vector machines. Multimed. Tools Appl. 2017, 76, 78037821. [CrossRef]

[7] Suk, M.; Prabhakaran, B. Real-time mobile facial expression recognition systemA case study. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, Columbus, OH, USA, 2427 June 2014; pp. 132137

[8] Torre, F.D.; Chu, W.-S.; Xiong, X.; Vicente, F.; Ding, X.; Cohn, J. IntraFace. In Proceedings of the IEEE International Conference on Automatic Face and Gesture Recognition, Ljubljana, Slovenia, 48 May 2015; pp. 18.

[9] Szwoch, M.; Pieni azek, P. Facial emotion recognition using depth data. In Proceedings of the 8th International Conference on Human System Interactions, Warsaw, Poland, 2527 June 2015; pp. 271277.

[10] Gunawan, A.A.S. Face expression detection on Kinect using active appearance model and fuzzy logic. Procedia Comput. Sci. 2015, 59, 268274.

[11] Polikovsky, S.; Kameda, Y.; Ohta, Y. Facial micro-expressions recognition using high speed camera and 3D-gradient descriptor. In Proceedings of the 3rd International Conference on Crime Detection and Prevention, London, UK, 3 December 2009; pp. 16

[12] Shen, P.; Wang, S.; Liu, Z. Facial expression recognition from infrared thermal videos. Intell. Auton. Syst. 2013, 12, 323333.

[13] Zhao, G.; Huang, X.; Taini, M.; Li, S.Z.; Pietikinen, M. Facial expression recognition from near-infrared videos. Image Vis. Comput. 2011, 29, 607619. [CrossRef]

[14] Wei, W.; Jia, Q.; Chen, G. Real-time facial expression recognition for affective computing based on Kinect. In Proceedings of the IEEE 11th Conference on Industrial Electronics and Applications, Hefei, China, 57 June 2016; pp. 161165.

[15] Paul Viola, Micheal J. Jones : Robust Real-Time Face Detection in International Journal of Computer Vision archive Volume 57 Issue 2, May 2004 Pages 137-154.

[16] Yin, L.; Wei, X.; Sun, Y.; Wang, J.; Rosato, M.J. A 3D facial Expression database for facial behavior research. In Proceedings of the International Conference on Automatic Face and Gesture Recognition, Southampton, UK, 1012 April 2006; pp. 211216.

[17] Lyons, M.J.; Akamatsu, S.; Kamachi, M.; Gyoba, J. Coding facial expressions with Gabor wave. In Proceedings of the IEEE International Conference on Automatic Face and Gesture Recognition, Nara, Japan, 1416 April 1998; pp. 200205

[18] Kahou, S.E.; Michalski, V.; Konda, K. Recurrent neural networks for emotion recognition in video. In Proceedings of the ACM on International Conference on Multimodal Interaction, Seattle, WA, USA, 913 November 2015; pp. 467474.

[19] Walecki, R.; Rudovic, O. Deep structured learning for facial expression intensity estimation. Image Vis. Comput. 2017, 259, 143154

[20] Benitez-Quiroz, C.F.; Srinivasan, R.; Martinez, A.M. EmotioNet: An accurate, real-time algorithm for the automatic annotation of a million facial expressions in the wild. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 26 June1 July 2016; pp. 55625570.

[21] Kolakowaska, A. A review of emotion recognition methods based on keystroke dynamics and mouse movements. In Proceedings of the 6th

International Conference on Human System Interaction, Gdansk, Poland, 68 June 2013; pp. 548555

[22] Walecki, R.; Rudovic, O.; Pavlovic, V.; Schuller, B.; Pantic, M. Deep structured learning for facial action unit intensity estimation. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 2126 July 2017; pp. 34053414.

[23] LeCun, Y.; Boser, B.; Denker, J.S.; Henderson, D.; Howard, R.E.; Jackel, L.D. Backpropagation applied to handwritten zip code recognition. Neural Comput. 1989, 1, 541551. [CrossRef]

[24] Ko, B.C.; Lee, E.J.; Nam, J.Y. Genetic algorithm based filter bank design for light convolutional neural network. Adv. Sci. Lett. 2016, 22, 23102313. [CrossRef]

[25] Breuer, R.; Kimmel, R. A deep learning perspective on the origin of facial expressions. arXiv 2017, arXiv:1705.01842 .