A Simple Multi-Feature Classification to Recognize Human Emotions in Images

Navin K Ipe

Abstract—The recognition of emotions via facial expressions is a complex process of piecing together various aspects of each facial feature. Since viewing a single facial feature in isolation may result in an inaccurate recognition of emotion, this paper attempts training neural networks to first identify specific facial features in isolation, and then use the general pattern of expressions on the face to identify the overall emotion. The technique presented is very basic, and can definitely be improved with more advanced techniques that incorporate time and context.

Index Terms—emotion recognition, image processing, neural network.

I. INTRODUCTION

Images are fed to a computer algorithm as a series of pixels that form a pattern. Once a human face is isolated from this pattern, the recognition of emotions from a picture requires the identification of specific facial features, extraction and reduction of those features and classifying the overall understanding of the features to recognize an emotion. Facial expressions could generally be described as motions or positions of facial muscles. Although many advanced methods of using Convolutional Neural Networks (CNN) have been developed, this paper attempts to find the simplest possible technique of classifying emotions, as a process of uncovering the fundamentals of emotion recognition. Emotion recognition techniques [1] [2] generally utilize concepts of optical flow, flow vectors, principal component analysis (PCA), hidden Markov models (HMM) etcetera to recognize emotions. However, the realities of how the human brain processes emotion recognition, appears to stem from a far more fundamental process that accounts for more information processing and understanding than a mere interpretation of facial features. The interpretation of emotion from emoticons, birds or other animals are one such example. Therefore, rather than merely creating a database of facial expressions, a decision was taken to allow neural networks to learn from a set of models of each facial feature to classify an emotion. The faces chosen for the experiment were taken from the JAFFE dataset [3] containing faces pre-classified (as a ground-truth) into seven emotions as shown in Fig. 1. Pre-processing techniques like histogram equalization, Otsu thresholding and erosion were applied to reduce features and highlight facial features that convey information about emotion.

Author acknowledges the support received from M S Ramaiah University of Applied Sciences, Bengaluru, India. N. K. Ipe is with the Department of Computer Science and Engineering, M S Ramaiah University of Applied Sciences, Bengaluru, India (e-mail: navinipe@gmail.com).

Manuscript received July 15, 2020

Fig. 1: Seven emotions

Section II presents similar research on emotion recognition, Sect. III presents the process by which emotion recognition was achieved, Sect. IV presents inferences derived from Sect. III and the paper concludes with Sect. V.

II. RELATED WORK

Prior work on classifying emotions, has explored various techniques like Bayesian networks, decision trees, Support Vector Machines (SVM), k-Nearest Neighbours (kNN), boosting and bagging [4]. Some earlier approaches have extracted geometric facial features, facial areas and texture of skin (furrows, wrinkles etc.), supplied facial features as inputs to a single backpropagation neural network and classified the results using data about the eyes and mouth [5]. Some researchers have used even Sobel filters to emphasize prominent edges and extracted features using pixel density [6]. Work by Konar and Chakraborty [7] has identified various methods of feature extraction, including facial, geometric model, appearance-based, voice features, gesture and posture-based features and even electroencephalogram-based methods. PCA, Independent Component Analysis (ICA) and evolutionary algorithms for feature reduction. Neural networks, SVM, Learning Vector Quantization (LVQ), Fuzzy methods, HMM, kNN and Naive Bayes methods for classification and various combinations of these approaches as multimodal methods. More recent methods of emotion recognition utilize CNN's, and are capable of achieving more than 95% accuracy [8] [9].

III. METHODOLOGY

The primary techniques considered for every stage of emotion recognition were PCA, SIFT (Scale Invariant Feature Transform), FACS (Facial Action Coding System), FAP (Facial Animation Parameters), ICA (Independent Component Analysis), FLD (Fisher's Linear Discriminant), HMM, SVM, Adaboost, Haar features, image thresholding and binarization. A decision was taken to automate all stages of the process and use the simplest technique possible, and this decision led

Fig. 2: The need for threshold choice

Fig. 3: Choice of threshold method

to good results, as shown in Sect. IV. PCA and ICA were not used since PCA fared poorly in initial trials. SVM was not used simply because a higher generalization ability was needed, since the facial features varied highly depending on the subject. Haar features were too crude to be capable of capturing subtleties in expression. SIFT was not necessary since the images had the faces centered already.

Figure 2 shows how the histogram varies for images taken under variations of illumination, so a hard-coded threshold value is insufficient. A few more trials (Fig. 3) demonstrated how (in Fig. 3a), an image with the appropriate Otsu binary thresholding is capable of capturing as much information as Fig. 3b which has a higher number of thresholds. Figure 3c is morphologically eroded. Although the second image appears to hold more information (furrows on forehead and more prominent nose furrows) than the first image, in reality, Fig. 3a possesses information that is more than sufficient to capture the emotional state of the subject. Thus, complex emotion recognition techniques were not necessary. A simple image thresholding and binarization was sufficient. It is interesting to note that capturing eye region was not necessary since eye movements caused corresponding movements in the eyebrow, which were more than sufficient for interpretation. Additionally, various eyes of each subject in the dataset were of varying shades which could have resulted in false positives (for example, a dark neutral eye being misinterpreted as a wide open eye). A few initial experiments were manually scrutinized in an image analysis tool by randomly displaying an image with the eyes hidden and even with eye information missing, and it was noticed that it was still possible to identify the emotion. After having performed this research on finding the simplest possible way to identify emotions using the least amount of information, it was decided that the eyebrow information, the nose furrow information and the mouth and chin involvement

information were sufficient to decide emotions.

A Self Organizing Map (SOM) was initially considered to automatically classify various eyebrow, mouth and nose furrows into classes and then use a combination of those class outputs to determine the emotion. However, since SOM is meant for unsupervised clustering, it would be unable to associate the clustered data with labels. So neural networks were attempted using a separate network for each feature. The issue arose when given the varied nature of eyebrow movement for any given emotion among the JAFFE subjects, predictions from the neural network were inaccurate. This obstacle led to the idea of training a separate neural network for each facial feature, which would interpret the emotion for that combination, as shown in Fig. 4. Each facial feature is identified under rectangular regions composed of s number of squares as shown in Fig. 5. Each facial feature may have a different value selected for s . The score value of each s is calculated as $v = 1$ if more than 50% pixels in s are black pixels. Else, $v = 0$. Scores fed to the neural network are calculated as shown in Eqs. 1 – 5. The neural networks in Fig. 4 are labeled “NN” for convenience, and are referred to as NN_1 to NN_5 in algorithm 1. The scores in Eqs. 1 – 5 are an aggregation of scores of each of the respective feature’s v , and the set of values are fed as an input to the neural network. The output of each neural network is a value E ranging from 1 to 7, corresponding to the seven emotions, as shown in Fig. 1. NN_6 is the final Emotion neural network that aggregates inputs from NN_1 to NN_5 .

$$S_{leftEyebrow} = score1 = \{v_i\}_{i=1}^{i=s_1} \quad (1)$$

$$S_{rightEyebrow} = score2 = \{v_i\}_{i=1}^{i=s_2} \quad (2)$$

$$S_{leftNoseFurrow} = score3 = \{v_i\}_{i=1}^{i=s_3} \quad (3)$$

$$S_{rightNoseFurrow} = score4 = \{v_i\}_{i=1}^{i=s_4} \quad (4)$$

$$S_{mouth} = score5 = \{v_i\}_{i=1}^{i=s_5} \quad (5)$$

Fig. 4: Prediction model

As shown in Fig. 5, since the faces were already centered, specific rectangular emotion capture regions were manually selected for examination. The eyebrow movements required a higher granularity to be captured, so they were divided

TABLE I: Results of emotion detection

Trial	MSE	Training images	Test images	No. of neurons	Time taken (s)	Accuracy (%)
1	4.66^{-26}	104	104	Left eyebrow: 60, Right eyebrow:60, Left nose furrow: 2, Right nose furrow: 2, Mouth: 24	147.72	100
2	4.08^{-30}	7	104	Left eyebrow: 60, Right eyebrow:60, Left nose furrow: 2, Right nose furrow: 2, Mouth: 24	158.2	25
3	1.02^{-28}	65	104	Left eyebrow: 60, Right eyebrow:60, Left nose furrow: 2, Right nose furrow: 2, Mouth: 24	303.77	67.30

Fig. 5: Granularity of the focus areas

Algorithm 1 Emotion Trainer

- 1: Load training images.
- 2: For each emotion, load corresponding image and calculate face score.
- 3: Train NN_1 with $S_{leftEyebrow}$ and E .
- 4: Train NN_2 with $S_{rightEyebrow}$ and E .
- 5: Train NN_3 with $S_{leftNoseFurrow}$ and E .
- 6: Train NN_4 with $S_{rightNoseFurrow}$ and E .
- 7: Train NN_5 with S_{mouth} and E .
- 8: For each emotion, load corresponding images, calculate scores 1 to 5 and train a radial basis function-based NN_6 with the scores as input and E as the output.
- 9: Load test images, calculate scores 1 to 5 and feed the scores to NN_6 to obtain the predicted emotion E .

into squares of pixel size 5. The nose furrow required low granularity, so 15 pixels were sufficient. The mouth needed a mere 10 pixel granularity. Since some rectangular regions extended beyond the face at the edges, those rectangles were ignored for computation (visible as the missing squares at the edges, above the eyebrow in Fig. 5). The model creation is automated by a series of neural networks. Each facial feature neural network was trained with seven emotion types: anger, disgust, fear, happiness, neutral, sadness and surprise. Each of these neural nets, when presented with a new image for classification, were able to predict the correct emotion only with a varying degree of accuracy for each facial feature. For example, for a happy image, the eyebrows would predict a high degree of neutral emotion, the nose furrows would predict disgust and the mouth would predict a high degree of happy expression. This gave the rise to the need for a separate emotion neural net which would be trained to recognize which combination of these facial expressions constitute a particular emotion, much like how humans use a combination of facial features to recognize an emotion. This observation

also highlights a very important step that is often ignored in image processing, which is that the interpretation of any image should be performed using all the components of detected patterns, in order to obtain the correct context. Five such neural networks as shown in Fig. 4, channeled their outputs into a radial basis function neural network which iteratively calculated the optimum number of hidden nodes required for minimizing the mean squared error (MSE).

IV. RESULTS

Images in the JAFFE dataset were manually examined for ambiguity in recognition of emotion. Ambiguous images were deleted and 104 images were retained. The results of the trials are shown in Table I. In the first trial, all 104 images were used for testing and training, which resulted in a 100% accuracy. Only one image of each emotion was chosen for training, resulting in a total of seven images that were selected manually for best conveying an emotion. Surprisingly, it gave a 25% accuracy. This shows that the model is capable of generalizing reasonably well and does not need to be trained with too many features. The third trial demonstrated how detection accuracy is dependent on the variety of training images.

Fig. 6: Mean squared error of trial 1

V. CONCLUSION

A division-of-labour technique for emotion recognition was attempted that focused on the simplest method of obtaining information from specific facial features fed to individual neural networks and the combination of the outputs was used as a unique pattern fed to an additional neural network that observed the overall interpretation of emotion. This method

was noted to require only information from the eyebrows, nose furrows, mouth and chin to be able to accurately judge the emotion. The images were already centered, so no extra effort was required for face detection and centering. The results show promise in the capability to recognize emotions from a few features of a single pose, and especially the results of training the neural network with just one image of each emotion that resulted in 25% accuracy. Future work could tackle varied granularities in feature extraction and an additional neural network for identifying the right image thresholds, based on histograms of various sections of the image.

REFERENCES

- [1] M. Z. Uddin and T.-S. Kim, "Facial expression recognition using independent component features and hidden markov model," *Emotion Recognition: A Pattern Analysis Approach*, p. 113, 2014. I
- [2] P. MS and K. N. GM, "Modified emotion recognition system to study the emotion cues through thermal facial analysis," *Biomedical Research (0970-938X)*, vol. 28, no. 20, 2017. I
- [3] M. J. Lyons, S. Akamatsu, M. Kamachi, J. Gyoba, and J. Budynek, "The japanese female facial expression (jaffe) database," in *Proceedings of third international conference on automatic face and gesture recognition*, 1998, pp. 14–16. I
- [4] N. Sebe, Y. Sun, E. Bakker, M. S. Lew, I. Cohen, and T. S. Huang, "Towards authentic emotion recognition," in *2004 IEEE International Conference on Systems, Man and Cybernetics (IEEE Cat. No. 04CH37583)*, vol. 1. IEEE, 2004, pp. 623–628. II
- [5] 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 *2013 12th Mexican International Conference on Artificial Intelligence*. IEEE, 2013, pp. 85–96. II
- [6] I. Perikos, E. Ziakopoulos, and I. Hatzilygeroudis, "Recognizing emotions from facial expressions using neural network," in *IFIP International Conference on Artificial Intelligence Applications and Innovations*. Springer, 2014, pp. 236–245. II
- [7] A. Konar and A. Chakraborty, *Emotion recognition: A pattern analysis approach*. John Wiley & Sons, 2015. II
- [8] L. Nwosu, H. Wang, J. Lu, I. Unwala, X. Yang, and T. Zhang, "Deep convolutional neural network for facial expression recognition using facial parts," in *2017 IEEE 15th Intl Conf on Dependable, Autonomic and Secure Computing, 15th Intl Conf on Pervasive Intelligence and Computing, 3rd Intl Conf on Big Data Intelligence and Computing and Cyber Science and Technology Congress (DASC/PiCom/DataCom/CyberSciTech)*. IEEE, 2017, pp. 1318–1321. II
- [9] H.-D. Nguyen, S. Yeom, G.-S. Lee, H.-J. Yang, I.-S. Na, and S.-H. Kim, "Facial emotion recognition using an ensemble of multi-level convolutional neural networks," *International Journal of Pattern Recognition and Artificial Intelligence*, vol. 33, no. 11, p. 1940015, 2019. II