Facial Emotions Recognition using Gabor Transform and Facial Animation Parameters with Neural Networks

Aditya Harit; Col J. C. Joshi; K. K. Gupta

Department of Electrical and Electronics Engineering, Birla Institute of Technology & Science, Pilani

kgupta@pilani.bits-pilani.ac.in

Abstract. The paper proposed an automatic facial emotion recognition algorithm which comprises of two main components: feature extraction and expression recognition. The algorithm uses a Gabor filter bank on fiducial points to find the facial expression features. The resulting magnitudes of Gabor transforms, along with 14 chosen FAPs (Facial Animation Parameters), compose the feature space. There are two stages: the training phase and the recognition phase. Firstly, for the present 6 different emotions, the system classifies all training expressions in 6 different classes (one for each emotion) in the training stage. In the recognition phase, it recognizes the emotion by applying the Gabor bank to a face image, then finds the fiducial points, and then feeds it to the trained neural architecture.

1. Introduction

Facial emotion recognition is a highly challenging area in machine vision due to the variations in facial gestures, illuminations, poses, background, etc. Different cultures in different parts of the world do not express same emotion in exactly the same way. Even human emotions are compounded by the fact that humans being often experience simultaneous emotions. For example, a person being upset with something may also show anger or fear. This makes it difficult for an automated recognition system to grasp the emotion accurately. Researchers have identified six basic emotions that all humans show [15]: anger, fear, disgust, happiness, surprise and sadness. Apart from these, there are other expressions which are "voluntary" expressions, which differ from culture to culture. Those are hard to distinguish, and not the focus of this paper.

Emotion recognition can take two main approaches: geometrical based and general techniques [1-2]. Geometrical techniques rely on ratios and measurements between various facial aspects such as eyes, nose, lips, etc. General techniques treat the whole face as a single object. In this paper, we describe an approach which uses both Gabor transform and Facial Action Coding System (FACS), and then uses supervised learning to train a neural network to recognize emotions.

2. Proposed Algorithm

The feature extraction is the backbone of the whole system. If good features are not extracted, the training phase will invariably fail. The feature extraction in Facial Expression Recognition (FER) problems is of two types: aggregated template matching structures and geometric feature based [5]. In aggregated setups, a feature vector is obtained after processing the image as a whole. No attention is paid to the individual components of a face. In geometric based setups, a dimensionality reducing technique such as PCA or ICA and MPANN are judiciously used to achieve a low dimensional

portrayal. In geometric feature dependent structures, primary face and fiducial points are first found in face images. The separations between fiducial points and the normalized representation of the primary face components are calculated to form a feature vector. These holistic approaches are usually computationally costlier than template based, but are less dependent upon changes in illumination, scale, size, orientation of subjects' head, and location of the face in an image. The work in this paper is a mixed approach. We first find a set of fiducial points, and then use them to extract FAPS [14] and then extract a set of Gabor wavelet coefficients at each point through the operation of convolution. After that, we feed the feature vector thus obtained in a two layered MPANN for training purposes, and the network trained is used for testing purposes later on.

3. Implementation

There are many algorithms for extracting fiducial points. The one we used to be developed by Kazemi and Sullivan [4]. In this approach, the facial image is first normalized by moving it to a coordinate system based on a current prediction of the points. Afterwards, the fiducial points are found to estimate an update vector for the shape parameters. This procedure is repeated many times until the error rate in the estimated and current points become negligible. With this, a total of 68 points are extracted.

Two operations are performed on these points: First, a Gabor filter bank of 6 orientations and 3 spatial frequencies is created. Then, Gabor transform is applied at each of the extracted points for all the specific filters present in the filter bank. This is done by taking convolution at a point (a, b) in the image. Such a Gabor transform applied at a point results in a complex Gabor coefficient of that point, containing both real and imaginary points. Since the phase of the Gabor varies greatly, while the magnitude remains relatively constant, we take the magnitude as the single feature of a point. When the whole filter bank is applied, we get 18 such magnitudes for each fiducial point in the image. If we take a maximum of 68 such fiducial points in the image, we get a vector of 18x68 = 1224 elements for each image.

Twenty-one of the extracted fiducial points are chosen for extracting relevant facial ratios. These are: 5 points for each eye-ball (left, right, top, bottom and center of the eye), 4 for the mouth, and 2 for each eyebrow, and 3 for the jaw Next, 14 relevant ratios are chosen using these points, as listed in the Table 1.

Table 1. Iterations and accumulated error.

Serial number	Description
1	Left outer eyeball to left outer eyebrow
2	Right outer eyeball to right outer eyebrow
3	Left inner eyeball to left inner eyebrow
4	Right inner eyeball to right inner eyebrow
5	Height of the left eyeball
6	Height of the right eyeball
7	Width of the left eyeball
8	Width of the right eyeball
9	Distance between center of two eyeballs
10	Distance between the eyebrows and the forehead.
11	Middle of the mouth to jaw
12	Width of the mouth
13	Middle of the upper lip to middle of the lower lip

14	Lower lip to jaw

4. Training the Neural Network

Once the features are extracted then neural network is design to train. A two-layered architecture was used. The input vector for each image was a vector of dimensionality 1238. Out of these, 1224 were from the Gabor bank, while the last 14 were from the face ratio. The features from the Gabor bank are indicative of a holistic approach, while the ones from the facial ratios indicate a geometric approach. For the hidden layer, a hyper-tangent function was used.
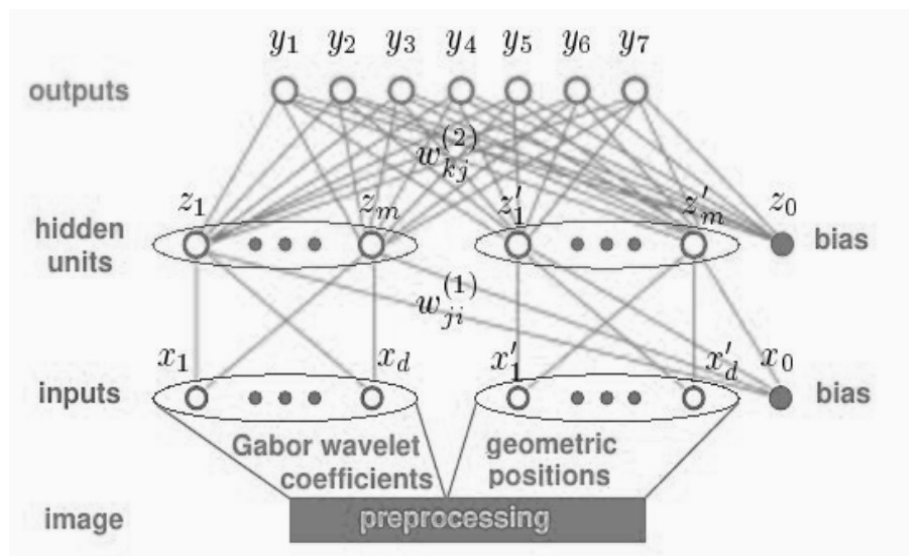

Fig 1. Architecture of neural network

As shown, the input was broken into two parts, which are fed to separate hidden layers. Note that there are no connections in the first layer among geometric and Gabor coefficient parts, since they are two separate pieces of information (holistic and geometric). The second layer then makes a decision depending upon the low dimensional set of features in the hidden nodes. Each output node is bound to a different emotion, so our system contains six output units (7 in the case where a neutral emotion is also present). When provided a test image, each output unit expresses the probability of correct emotion, which is related to that output unit. When training the neural network, RPROP was used, as it provides faster response than simple backpropagation.

5. Testing the proposed Model

The model was tested on two datasets: The Yale Face Database A, provided by Computer Science and Engineering U.C. San Diego. It contains The 165 grayscale images in GIF format of 15 individuals. There are 11 images per subject, one per different facial expression or configuration: center--light, w/glasses, happy, left--light, w/no glasses, normal, right--light, sad, sleepy, surprised, and wink. Since the task is for gesture recognition so the images are removed which had no categorization from this dataset. This left 90 images in the dataset.

The Japanese Female Facial Expression (JAFFE) Database, provided by Michael Lyons, Miyuki Kamachi, and Jiro Gyoba of Psychology Department in Kyushu University. This database contains 213 images of 7 facial expressions (6 basic facial expressions + 1 neutral) posed by 10 Japanese female models. Each image has been rated on 6 emotion adjectives by 60 Japanese subjects.

The testing of results was done in various ways. A general methodology was adopted for all 3 ways: First partition the data into two sets, one for training, and another for testing randomly. Then train the two-layer perceptron on the training data, and tests its performance on the testing data. This is done multiple times, and the final network is the average of all the resulting networks. The training was usually stopped after 250 epochs.

When tested on Yale database, the algorithm gave a result of 67% accuracy (33% error) in emotion recognition rate on the test data. This was when the number neuron in the hidden layer was 7. As the number of neurons increased, over-fitting was observed, increasing sharply after it rose above 10 units, giving an accuracy of less than 25% for 11 units. This is not surprising, as the number of images used was very low and the Yale database has greater ethnic diversity than the JAFFE database.

When tested on the JAFFE database however, the algorithm gave an accuracy of 81%, with similar trends in overfitting as for Yale database [Fig 3]. This is a good result; as conventional databases typically have around 500 or more images in them. The similar ethnicity of people in JAFFE databases also probably contributed to this effect.
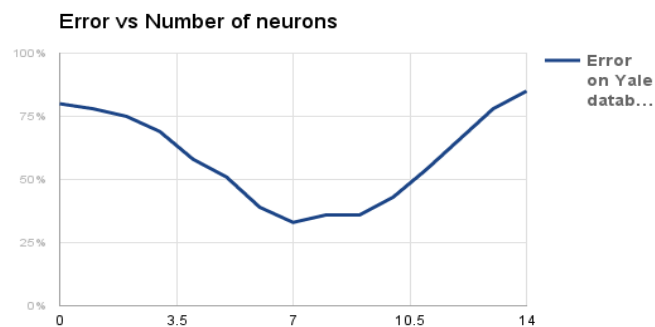
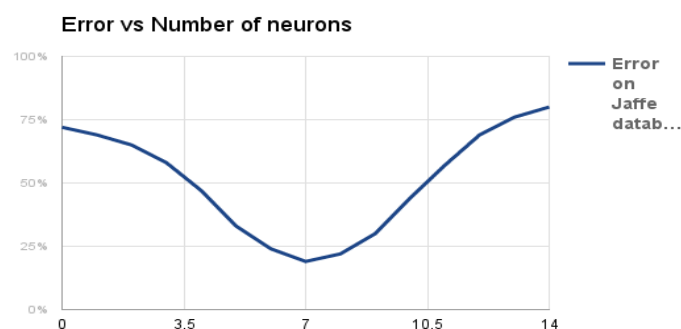

Figure 2. Error on Yale database

Figure 3. Error on Jaffe database

6. Conclusions and Future work
In this paper, facial emotion recognition is investigated. A hybrid model based on both geometric and holistic approaches is developed, which performed admirably considering the low number of images used. The accuracy could have been improved if automatic fiducial point marker improves. The ethnicity of subjects in YALE database caused problems, as the system could not effectively learn

from the a relatively small sample of pictures. This is clearly shown as the JAFFE database yields a result of about 20% more accuracy. A single emotion was easily detected with a very high accuracy, of over 90%. This shows that although the system is not very adept at recognising different emotions, it can easily serve as a second layer of a system where only an approximate guess of emotion is made. Less than 5 units in the hidden layer do not give a satisfactory result, as the dimensionality of the input is not able to be conveyed by such a small number of nodes. Increasing the number of hidden nodes is not always beneficial, and leads to overfitting More work is currently being done on this, integrating new approaches.

References
[1]   M. Suma, N. Sugie, and K. Fujimora. A preliminary note on pattern recognition of human emotional expression. In Proceedings of the 4th International Joint Conference on Pattern Recognition, pages 408–410, 1978.
[2]   K. Mase. Recognition of facial expression from optical flow. IEICE Transactions E, 74(10):3473–3483, 1991.
[3]   Z. Zhang et al. Comparison Between Geometry Based and Gabor Wavelets Based Facial Expression Recognition Using Multi-Layer Perceptron.
[4]   Vahid Kazemi and Josephine Sullivan KTH. One Millisecond Face Alignment with an Ensemble of Regression Trees. Royal Institute of Technology Computer Vision and Active Perception Lab.
[5]   T. Hastie, R. Tibshirani, and J. H. Friedman. The elements of statistical learning: data mining, inference, and prediction. New York: Springer--Verlag, 2001
[6]   Christian Igel and Michael Hüsken. Empirical Evaluation of the Improved Rprop Learning Algorithm. Neurocomputing 50:105-123, 2003.
[7]   Lee, Tai S. Image Representation Using 2D Gabor wavelets. IEEE Transactions on Pattern Analysis and Machine Intelligence, October 1996.
[8]   M. Slavkovic et al. Face Recognition using Gabor filters, PCA, and Neural Networks 2013 20th International Conference on Systems, Signals and Image Processing (IWSSIP)
[9]   C. Bishop. Neural Networks for Pattern Recognition. Clarendon Press, Oxford, 1995.
[10]   PJ Werbos. Backpropagation through time: what it does and how to do it. Proceedings of the IEEE, 1990.
[11]   JR Movellan. Tutorial on Gabor filters. Open Source Document, 2002 - cs.olemiss.edu
[12]   L Shen, L Bai. A review on Gabor wavelets for face recognition, Pattern analysis and applications, 2006.
[13]   A. Hyvarinen, J. Hurri, P.O. Hoyer. Natural Image Statistics A probabilistic approach to early computational vision. Springer, February, 2009.
[14]   A. Raouzaiou, S. Ioannou, K. Karpouzi,N. Tsapatsoulis, S. Kollias, R. Cowie. An Intelligent Scheme for Facial Gesture Recognition. Springer, 2003.
[15]   P. Ekman and W. Friesen. The Facial Action Coding System. Consulting Psychologists Press, San Francisco, CA, 1975