Facial Expression Recognition in Static Images

Group 24 - Bhavya Patwa, Karan Patel, Ratnesh Shah, Yash Kotadia
School of Engineering and Applied Science
Ahmedabad University

***Abstract*—Behaviours, actions, poses, facial expressions and speech; these are considered as a medium which conveys human emotions. Facial expression analysis is rapidly gaining intense interest in the research field. It is an interesting and challenging problem and have important applications in many areas such as human-computer interaction and data-driven animation. In this project we present an automated system to recognize facial expressions from static facial images. Thus a multiclass based solution combined with image processing is used in classifying universal emotions such as Happiness, Sadness, Anger, Disgust and Surprise.**

I. Introduction

An emotion is a mental and physiological state which is subjective and private; it involves a lot of behaviors, actions, thoughts and feelings. Many factors contribute in conveying emotions of an individual such as pose, speech, facial expression, behaviour and actions are some of them. From this above mentioned factors facial expression plays an important role since they can be easily visualized. In communicating with others humans can recognize emotions of other humans with high level of accuracy. Expression recognition system consists of three modules, face detection, facial feature extraction and facial expression recognition. First, a static image is taken then face detection algorithm is applied. After the face is detected, image processing based feature point extraction based method is used to extract a set selected feature points. Finally, a set of values obtained after processing those feature points are given as input to the trained multiclass SVM model to recognize the emotion.

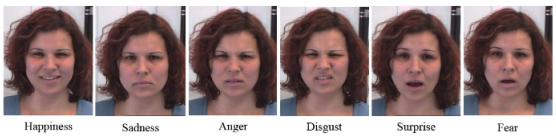

Fig. 1: Emotions

II. Motivation

Significant debate has risen in the past regarding the emotions portrayed in the world famous masterpiece of Mona Lisa (Figure 2). British weekly 'New Scientist' has stated that she is in fact a blend of many different emotions, 83% happy, 9% disgusted,6% fearful and 2% angry. Modern drifts in emotion transferring to 3D avtaars, games and nonverbal communication interpretations have made this area attractive.

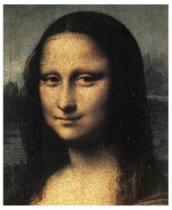

Fig. 2: Mona lisa

III. Data Set

The Data Set which is used is Cohn-Kanade(CK+) data set and it contains behavior of 210 adults and it was recorded using two hardware synchronized Panasonic AG-7500 cameras. This database was used for facial expression recognition in six basic facial expression classes (anger, disgust, fear, happiness, sadness, and surprise). This database consists of 593 sequences from 123 subjects. The image sequence varies in duration (i.e., seven to 60 frames), and incorporates the onset (which is also the neutral face) to peak formation of the facial expression. Participants were instructed by an experimenter to perform a series of 23 facial displays; these included single action units and combinations of action units. Each display began and ended in a neutral face with any exceptions noted. Only 327 of the 593 sequences have a given emotional class. This is because these are the only ones that fit the prototypical definition.

We have filtered out the dataset such that we have the first We have performed face detection and feature extraction on a data set of 654 Images containing faces with 8 different emotions:

Expression	Images
Anger	45
Contempt	18
Disgust	59
Fear	25
Happy	69
Neutral	327
Sadness	28
Surprise	83

IV. Model Description

The process of facial expression recognition can be broadly categorized into three steps:
A. Face Detection
B. Facial Feature Extraction
C. Emotion Classification.

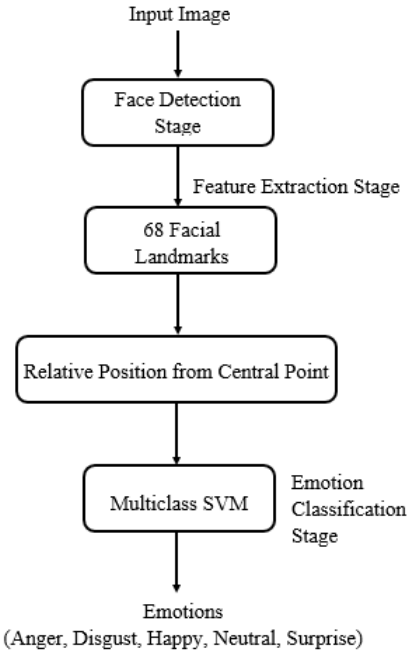

Fig. 3: Model

We initially divide the dataset into two parts which contains 80% images for training and 20% for testing.

A. Face Detection

Given an image, detecting a human face is difficult task because of large possible variations of the face. The different sizes, angles and poses a human face have within the image can cause the variation. Different imaging conditions such as illumination and occlusions also affect the image. Also, the presence of spectacles, beard, hair and make up have an effect in the facial appearance.

The methods for face detection can be broadly classified into four sections:
1. Knowledge based approach
2. Feature Invariant approach
3. Template based approach
4. Appearance based approach

Knowledge-based approach: It is based on rules derived from the knowledge on the face geometry.

We have used Haar feature-based cascade classifiers for detection of frontal faces in the image. Initially, the algorithm needs a lot of positive images (images of faces) and negative images (images without faces) to train the classifier. Then we need to extract features from it. They are just like our convolutional kernel. Each feature is a single value obtained by subtracting sum of pixels under white rectangle from sum of pixels under black rectangle.
For each feature, it finds the best threshold which will classify the faces to positive and negative. But obviously, there will be errors or misclassifications. We select the features with minimum error rate, which means they are the features that best classifies the face and non-face images.

We try to detect the face using the four different types of haarcascade files which are
1. frontal_face_default.xml
2. frontal_face_alt2.xml
3. frontal_face_alt.xml
4. frontal_face_alt_tree.xml

We then cut the frame such that only the face part remains and then resize the image so that all the images are of the same size.

B. Facial Feature Extraction

We get the 68 landmark points then the first thing to do is find ways to transform these landmark points overlaid on your face into features to feed the classifier. Features are little bits of information that describe the object or object state that we are trying to divide into categories.

If, for example, we would extract eye colour and number of freckles on each face, feed it to the classifier, and then expect it to be able to predict what emotion is expressed, we would not get far. However, the facial landmarks from the same image material describe the position of all the “moving parts” of the depicted face, the things you use to express an emotion. This is certainly useful information!

We extract the coordinates of all facial landmark points. These coordinates are the first collection of features. We saw that the coordinates change as the face moves to different parts of the frame. The person in the image could be expressing the same emotion in the top left of an image as in the bottom right of another image, but the resulting coordinate matrix would express different numerical ranges. However, the relationships between the coordinates will be similar in both matrices so some information is present in a location invariant form, meaning it is the same no matter where in the picture my face is.

So we tried the most straightforward way to remove numerical differences originating from faces in different places of the image would be normalising the coordinates between 0 and 1. This is easily done by:

$$X_{norm} = \frac{X - X_{min}}{X_{max} - X_{min}}$$

However, there is a problem with this approach because it fits the entire face in a square with both axes ranging from 0 to 1. Imagine one face with its eyebrows up high and mouth open, the person could be surprised. Now imagine an angry face with eyebrows down and mouth closed. If we normalise the landmark points on both faces from 0-1 and put them next to each other we might see two very similar faces. Because both distinguishing features lie at the edges of the face, normalising will push both back into a very similar shape. The faces will end up looking very similar.

So, now to get away from this problem we find the positions of all points relative to each other. To do this we calculate

the mean of both axes, which results in the point coordinates of the sort-of "centre of gravity" of all face landmarks. We can then get the position of all points relative to this central point. There is one last thing to note. Faces may be tilted, which might confuse the classifier. We can correct for this rotation by assuming that the bridge of the nose in most people is more or less straight, and offset all calculated angles by the angle of the nose bridge. This rotates the entire vector array so that tilted faces become similar to non-tilted faces with the same expression.

Now, we use this coordinates to train the Support Vector Machine. By doing so we give a holistic approach of expression recognition which works well even for input image not present in the training dataset. Hence our approach is highly suitable for real time emotion recognition.

C. Multiclass SVM for emotion Classification

The generated feature points of the training data we train the Support vector machine using different types of kernels to obtain the hyperplanes which separates all the emotions. A SVM is a binary classifier. The class labels can only take two values: ± 1. Several different schemes can be applied to the basic SVM algorithm to handle the K-class pattern classification problem. We have used one vs all approach for performing multiclass classification using SVM.

Why did we use SVM?

- Effective in high dimensional spaces.
- Still effective in cases where number of dimensions is greater than the number of samples.
- Uses a subset of training points in the decision function (called support vectors), so it is also memory efficient.
- Versatile: different Kernel functions can be specified for the decision function. Common kernels are provided, but it is also possible to specify custom kernels.

Using k one-vs-all classifiers is the simplest approach, and it does give reasonable results. K classifiers will be constructed, one for each class. The K^{th} classifier will be trained to classify the training data of class k against all other training data. The decision function for each of the classifier will be combined to give the final classification decision on the K-class classification problem,

$$f(x) = argmax_k \sum_{i=1}^{l} \lambda_i^k y_i K^k(x_i, x) + b^k \qquad (1)$$

The total number of classifiers for a K-class problem will then be K(K-1)/2. The training data for each classifier is a subset of the available training data, and it will only contain the data for the two involved classes. The data will be reliable accordingly, i.e. one will be labelled as +1 while the other as -1. These classifiers will then be combined with some voting scheme to give the final classification results, such as majority voting or pairwise coupling.

V. Results

We obtain an accuracy of 84.1%, when classifying the data into 8 emotions. We then reduce the set to 5 emotions (leaving out contempt,fear and sadness), because the three categories had very few images and these approach gives 91.10% accuracy a lot better than previous results.

The below table describes the accuracy obtained by using different kernels:

Kernel	Mean Accuracy
linear	91.10%
polynomial	89.40%
rbf	57.20 %

We also test our model on our class data set the following are the few results we obtained:

A. Output on Static Images

Fig. 4: Input of Static Image with Happy face

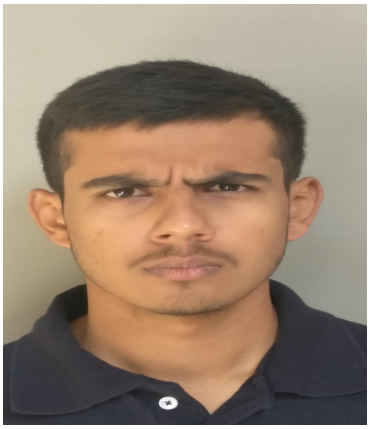

Fig. 5: Input of Static Image with Anger face

Fig. 6: Output of Both Static Image

B. Live Video

Fig. 7: Live video test with Happy Expression

Fig. 8: Live video test with Surprise Expression

C. Probability Bar-Graphs for Each Expression

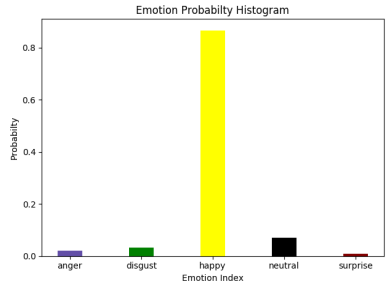

Fig. 9: Output for Fig 2(Monalisa Image)

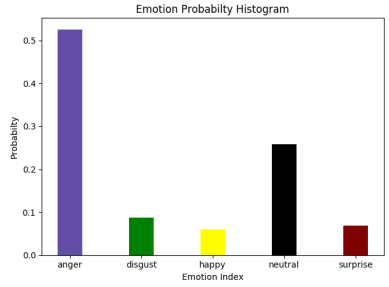

Fig. 10: Output for Fig 5

D. Confusion Matrix

We have generated confusion matrices of SVM using different kernels linear, polynomial and RBF.

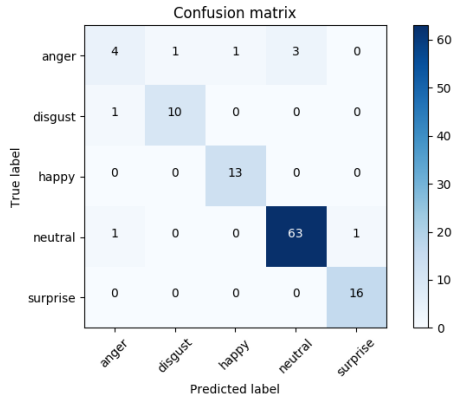

Fig. 11: Confusion Matrix of Linear Kernel

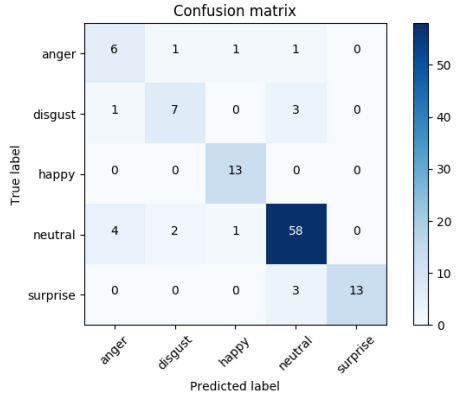

Fig. 12: Confusion Matrix of Polynomial Kernel

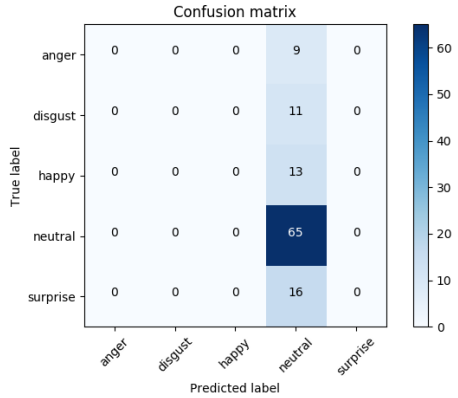

Fig. 13: Confusion Matrix of RBF Kernel

VI. Future Work

We have implemented SVM for multiclass classification, the training we use is still quite small in machine learning terms, containing about 1000 images spread over 8 categories. Further we will try to improve the efficiency by using deep neural network approach and currently we are classifying only into 5 emotions so we will try to accurately classify all major 8 emotions in real time scenario.

References

[1] Donato, G., Bartlett,M., Hager, J., Ekman, P., Sejnowski, T.: Classifying facial actions. IEEE Trans. Pattern Anal. Mach. Intell. 974–989 (1999)

[2] Fukui, K., Yamaguchi, O.: Facial feature point extraction method based on combination of shape extraction and pattern matching. Syst. Comput. Jpn. 29(6), 49–58 (1998)

[3] http://www.paulvangent.com/2016/08/05/emotion-recognition-using-facial-landmarks

[4] www.researchgate.net/publication/227031714 Facial Expression Recog.

[5] www.cs.utah.edu/widanaga/papers/Widanagamaachchi.2009.thesis.pdf

[6] Ping Du, Yankun Zhang, Chongqing Liu, Inst. of Image Processing and Pattern Recognition, Shanghai Jiao Tong University. "Face Recognition using Multi-class SVM", The 5th Asian Conference on Computer Vision, 23–25 January 2002, Melbourne, Australia.