

# Emotion Detection Through Facial Feature Analysis

Ritwick Chaudhary, Vishal B. Bhavani, Yash Sanghvi

**Abstract**—In the project, we propose to build an algorithm to detect the emotion present in face containing images. The dataset used contains faces for seven emotions namely (1) anger, (2) contempt, (3) disgust, (4) fear, (5) happy, (6) sadness and (7) surprise. We use Viola Jones cascade object detector to find out faces in the image and then the histogram of gradient (HoG) and LDA features are computed. Then using nearest neighbor and one-vs-all SVM techniques, we tried various features to detect the correct emotions and obtained 82% accuracy in the best method.

## I. DESCRIPTION AND RESULTS

We tried four different combinations of features and classification and obtained accuracy of each method using 10 Fold Cross Validation. For each of these methods, we use **Viola Jones object cascade detector** to obtain the bounding box for the face. The corresponding bounding box is then resized to  $64 \times 64$  pixels for uniformity across all test and cross validation points. The following section explains the further description of classification techniques used along with the training accuracies and confusion matrix.

- 1) **One-vs-all SVM on Raw Pixel Data:** In this technique we use the raw pixel data after normalising to (0,1) as the 4096 dimensional feature space (as  $64 \times 64 = 4096$ ) and train a one-vs-all SVM using the training data. The accuracy obtained in 10 fold cross validation was **76.25%**. The following is the confusion matrix:

$$\begin{bmatrix} 29 & 2 & 7 & 3 & 1 & 10 & 1 \\ 3 & 9 & 2 & 4 & 4 & 1 & 4 \\ 5 & 1 & 47 & 0 & 2 & 1 & 0 \\ 0 & 1 & 0 & 14 & 0 & 0 & 3 \\ 0 & 0 & 0 & 0 & 61 & 0 & 0 \\ 6 & 3 & 2 & 1 & 0 & 11 & 2 \\ 2 & 0 & 0 & 2 & 0 & 3 & 73 \end{bmatrix}$$

- 2) **Nearest Mean on LDA Feature Space:** In this technique, we firstly converted the raw pixels to a basis obtained by PCA and then applied LDA to obtain a  $|C| - 1$  dimensional feature space where  $|C|$  represents the number of classes. Then the test data was converted to  $|C| - 1$  space and the nearest mean in the projected space to classify the points. The accuracy obtained in 10 fold cross validation was **12.5%**.

TABLE I  
ACCURACY OF DIFFERENT TECHNIQUES

Method	Accuracy
One-vs-all SVM with Raw Pixel Data	76.25%
Nearest Neighbour on LDA Feature Space	12.5%
SVM on LDA Feature Space	80.31%
SVM on LDA + HoG Feature Space	82%

$$\begin{bmatrix} 11 & 6 & 3 & 5 & 10 & 6 & 14 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 3 & 3 & 0 & 2 & 16 & 2 & 13 \\ 4 & 0 & 6 & 0 & 4 & 0 & 1 \\ 17 & 3 & 18 & 7 & 8 & 6 & 16 \\ 1 & 0 & 14 & 3 & 5 & 4 & 21 \\ 9 & 5 & 17 & 8 & 24 & 8 & 17 \end{bmatrix}$$

- 3) **SVM on LDA Feature Space:** This technique is similar to previous one except that here we train a one-vs-all SVM to classify rather than the nearest neighbour rule. The accuracy obtained on 10 fold cross validation was **80.31%**.

$$\begin{bmatrix} 31 & 2 & 1 & 2 & 0 & 7 & 1 \\ 2 & 8 & 0 & 4 & 1 & 2 & 1 \\ 1 & 0 & 54 & 0 & 2 & 1 & 0 \\ 0 & 2 & 0 & 14 & 0 & 0 & 3 \\ 0 & 1 & 0 & 3 & 64 & 0 & 0 \\ 10 & 0 & 2 & 1 & 0 & 13 & 3 \\ 1 & 4 & 1 & 1 & 1 & 3 & 73 \end{bmatrix}$$

- 4) **SVM on LDA, HoG Feature Space:** We calculate the Histogram of gradients (HoG) feature space and concatenate it to the LDA feature space obtained in the previous technique. The overall accuracy obtained from 10 fold cross validation was **82%**.

$$\begin{bmatrix} 33 & 2 & 3 & 2 & 0 & 6 & 1 \\ 2 & 11 & 0 & 4 & 1 & 2 & 3 \\ 4 & 0 & 52 & 0 & 1 & 1 & 1 \\ 0 & 2 & 0 & 13 & 0 & 0 & 2 \\ 0 & 1 & 0 & 2 & 65 & 0 & 0 \\ 5 & 1 & 1 & 0 & 0 & 16 & 1 \\ 0 & 0 & 1 & 4 & 1 & 1 & 75 \end{bmatrix}$$

## II. OBSERVATIONS

- 1) We achieved marginal improvement over the original project in the LDA+HoG method. In the original method, they use LDA method to distinguish among the easier to distinguish classes and if predictions are none or more than one they use HoG features.
- 2) The HoG + LDA(4) and LDA(3) method performed better than training the SVM over raw pixel data. A possible reason is that these features are efficient at

removing redundant information not related to facial features.

- 3) Class 2 i.e. contempt was the most misclassified emotion throughout all the techniques often misclassified as fear, disgust or happy.
- 4) Though the LDA feature space was supposed to separate out the classes the most however, the LDA feature space didn't have the classes separated sufficiently as the accuracy by nearest neighbour method was very less i.e 12%.

### III. REFERENCES

- 1) Emotion Detection Through Facial Feature Recognition, James Pao, EE368/CS232 Course Projects
- 2) Lucey, P., Cohn, J.F., Kanade, T., Saragih, J. Ambadar, Z. The Extended Cohn-Kanade Dataset (CK+): A complete dataset for action unit and emotion-specified expression. IEEE Computer Society Conference CVPRW (2010)