Facial Expression Recognition

Andrew Zhai, Mengyi Wo, Raphael Townshend, Paul Ruan

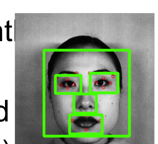
Problem Description

Recognizing emotions a human expresses is a useful problem to solve as it enables much more dynamic and social forms of human-computer interaction. Though both audio and video cues can be used to accomplish this task, this project focused on using single facial expressions images. Computer Vision techniques were used to attempt to recognize which one of six facial expressions were being displayed given static images of human subjects. The six facial expressions we considered were joy, sorrow, disgust, anger, fear, and surprise. Two separate techniques were attempted to solve this problem. Both an official, widely-used dataset, and a self-designed dataset were used.

Techniques - Points of Interest

Our initial idea is to extract interest points from faces and derive features from these points to train an SVM (inspired by [4]). To achieve the extraction goal, we first worked on extracting the following parts from a picture of a person: face, mouth, left eye, right eye, and nose. We trained multiple haar cascade classifiers, each corresponding to one of the parts were are trying to extract. Each classifier would return a rectangular area of the location that has the most probable match. The part extraction purely with the Haar cascade classifiers however was very poor, not achieving the robustness we desired. To achieve better results, we combined the haar cascade classifiers with information about the spatial information of an average human's face as follows:

- After finding the location of the face, run the other classifiers on the face portion only.
- Run the left/right eye classifier on the left/right portion of the face
- Run the nose classifier on the middle portion of the face
- Run the mouth classifier on the bottom portion of the face
- If only one of the eyes were found, use one eye's location to guess the other's.
- If only the nose or only the mouth is found, use one's location to find guess the other's

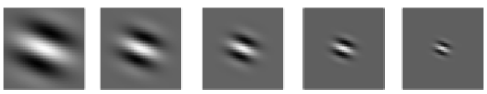


Although we detected the nose, we noticed that the nose did not significant change for the 6 different facial expressions and as a result we did not extract interest points on the nose. We did extract interest points for the mouth, eyes, and eyebrows (we used the location of the eyes to guess the location of the eyebrows). To extract these interest points, we used a combination of Harris corner detection and Canny edge detection. Canny was used to get the outline of the various facial parts we used to extract points. For the corners of the mouth and eyes, we used Harris corner detection because it was more robust.

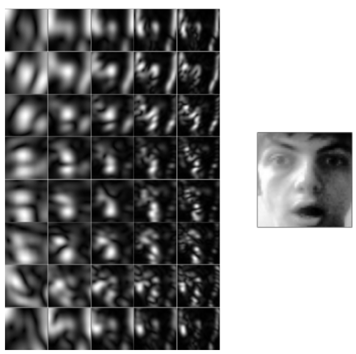
Although our initial plan was to train an emotion classifier based on a normalized set of these interest points, this technique was not used in our final emotion detector because the extraction could not handle non-ideal conditions such as lighting variations which was present in our BSAF dataset (see photos to the right).

Techniques - Banks of Gabor Filters

After studying some successes in facial expression recognition [1, 2, 6], we learned that responses to Gabor filters, Gaussians multiplied by a harmonic function, are very effective features for this task. Following the methods in these papers, we generated a bank of Gabor filters with different spatial frequencies and orientations. Features for each face were obtained by simply applying the bank of filters; the magnitude of the responses were then concatenated together into a feature vector. These features were then used to train six linear SVMs, one for each expression. At classification time, the prediction was based on the SVM with furthest distance from the margin, or the strongest confidence in the expression.



Example of Gabor filters with varying spatial frequency.

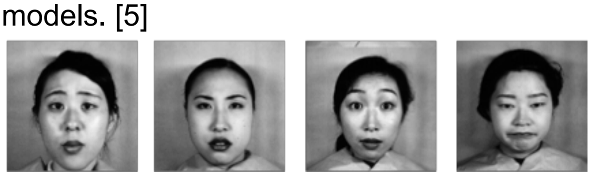


Magnitudes of Gabor filter responses along with source image.

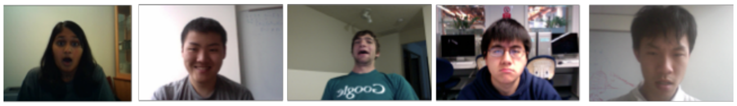
We preprocessed each image before extracting our features. First we detect a region containing a face. Then we crop the region to remove hair and background. We did this by trimming the edges (eg 10% of the height off the top). We also tried cropping an image based on interest points. We detected eye centers and used the distance between them to cut out the face based on metrics of a face model described in the paper by Deng et al. [6] which gave us slightly better results. The cropped images were resized to 48-by-48 pixel images.

Datasets

• The Japanese Female Facial Expression (JAFFE) dataset contains 213 images of 7 facial expressions (6 basic facial expressions and 1 neutral expression) posed by 10 Japanese female models. [5]



• The Berkeley Students And Friends (BSAF) dataset, collected by ourselves, contains 329 images of 7 facial expressions posed by 47 subjects. [7]



Both datasets contain frontal view photos. However, the JAFFE dataset has uniform lighting and position, while the BSAF dataset has variable lighting, background, and positions.

Results

We trained SVMs using responses to the bank of Gabor filters as features and tested each dataset separately. Datasets were divided roughly 75% into training and 25% into testing.

We achieved decent results on the JAFFE dataset. Overall, 37 out of 46 facial expressions were classified correctly giving us about 80.43% accuracy. Our test set turns out to have a lot more joys, which we seem to do very well on, than other expressions which may have skewed our accuracy a little but predictions of other expressions are not bad either. Below is the confusion table. The row labels are the ground truth, what the subject was asked to perform, and the column labels are the predictions. Each cell specifies how many of the true expression in the test set was classified as each of the six expressions.

	Joy	Sorrow	Disgust	Anger	Surprise	Fear
Joy	12	0	0	0	0	0
Sorrow	1	6	0	0	0	2
Disgust	0	0	2	3	0	1
Anger	0	0	3	6	1	0
Surprise	0	0	0	0	6	0
Fear	0	0	0	2	0	5

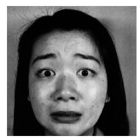
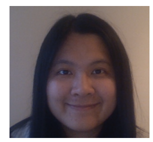
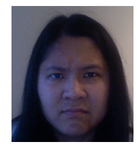
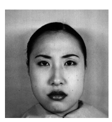
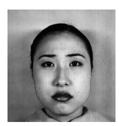
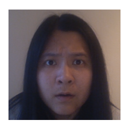
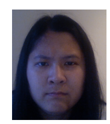
However, our accuracy on our BSAF dataset is not as good. 34 out of 61 facial expressions were classified correctly giving us about 55.74% accuracy. Below is the confusion table.

	Joy	Sorrow	Disgust	Anger	Surprise	Fear
Joy	10	0	0	0	0	0
Sorrow	0	3	1	4	1	1
Disgust	1	1	5	1	1	1
Anger	0	0	3	6	1	0
Surprise	1	1	0	0	7	1
Fear	0	1	1	1	5	3

The uniformity of the images may be a reason for the difference in accuracy. The JAFFE dataset has more uniform lighting and the faces are in about the same location and positioned upright, front-facing. The BSAF dataset, however, has varying lighting between images as well as within single images; for some images, half of the image is in light while the other half in shadows. Some subjects of the BSAF also rotate or lean their head significantly when expressing emotions. These differences can cause problems because important areas eyes and mouths would be in different locations, affecting the features we are using. Better preprocessing may help.

Another possibility for the difference is how well the subjects performed the expressions. Overall, through manual inspection, subjects in the JAFFE dataset seem to perform each expression more distinctively than the subjects in the BSAF dataset. As subjects ourselves, we found it difficult to force ourselves to show certain emotions like sorrow and fear.

Examples of predictions (a friend was asked to give her predictions too)

Correct				
Truth: Machine: A Human:	Surprise Surprise Surprise	Fear Fear Fear	Joy Joy Joy	Disgust Disgust Anger
Incorrect				
Truth: Machine: A Human:	Disgust Anger Sorrow	Sorrow Fear Sorrow	Fear Surprise Fear	Sorrow Anger Anger

References

1. M. S. Bartlett, G. Littlewort, I. Fasel, J. R. Movellan. Real Time Face Detection and Facial Expression Recognition: Development and Applications to Human Computer Interaction.
2. H. Seyedarabi, S. Fazli, R. Afrouzian. A Fast Algorithm For Recognition Of Basic Facial Expressions With Gabor Filter Bank
3. V. Bettadapura. Face Expression Recognition and Analysis: The State of the Art
4. A. Bajpai, K. Chadha. Real-time Facial Emotion Detection using Support Vector Machines
5. Michael J. Lyons, Shigeru Akamatsu, Miyuki Kamachi, Jiro Gyoba. Coding Facial Expressions with Gabor Wavelets
6. Deng, H.B., Jin, L.W., Zhen, L.X., Huang, J.C. A New Facial Expression Recognition Method Based on Local Gabor Filter Bank and PCA plus LDA
7. Contributors to the BSAF