Facial Expression Recognition

Manlin Xu, Xiangyu Li

1. Introduction

Major component of human communication are facial expressions which constitute around 55 percent of total communicated message. We use facial expressions not only to express our emotions, but also to provide important communicative cues during social interaction, such as our level of interest, our desire to take a speaking turn and continuous feedback signaling understanding of the information conveyed.

There are many ways that humans display their emotions. The most natural way to display emotions is using facial expressions. In the past 20 years there has been much research on recognizing emotion through facial expressions. This reasearch was pioneered by Ekman and Friesen [4] who started their work from the psychcology perspective. In the early 1990s the engineering community started to use these results to construct automatic methods of recognizing emotions from facial expressions in images or video [5, 6, 8, 7, 1] . Work on recognition of emotions from voice and video has been recently suggested and shown to work by Chen [1], Chen et al. [2], and DeSilva et al [3].

In this project, facial expression analysis refers to computer systems that attempt to automatically analyze and recognize facial motions and facial feature changes from visual information. Sometimes the facial expression analysis has been confused with emotion analysis in the computer vision domain. For emotion analysis, higher level knowledge is required.

The basic idea behind facial expression recognition is that everyone will share some face features when they have same emotion and make the corresponding faces. So we can extract these features and classify facial expressions based on these features with different classifying algorithm. In our project, we choose SVM algorithm to recognize facial expression because SVM as a static classifier has a higher accuracy.

2. Dataset

The database we are using 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. In this project, we only select 4 facial expressions (happy, sad, angry and neutral). Besides of that, the data set also contains several videos of our classmates, each of which includes about 1800 frames of 4 facial expressions(3 basic facial expressions + 1 neutral).

Labeling was done manually but unfortunately, there was no standard rule for grouping up similar facial expression which resulted in the problem of label inconsistency. Therefore, we resolve this issue by considering only the 4 most common facial expressions. Due to the relatively large size of dataset, limited computing power, and time constraints for our experiments, we made a smaller dataset by randomly selecting 1240 out of 5362 frames. For the whole dataset, including Japanese female models and the video of Lu Han and Holly, we split the data into 1006 and 234 frames for a training set and test set, respectively, while ensuring that frames from the same scene are not split over both the training and test set. Table 1 shows the statistics of our dataset.

Facial Expression	Training Frames	Test Frames
Happy	225	59
Sad	264	68
Angry	288	57
Neutral	229	50
Total	1006	234

Table 1. statistics in our 1240 frame dataset

3. Face detection and feature extraction based on Haar-like feature

We used haar-like features technology to detect face and extract features. We used haar-like features technology to detect face and extract features. Each Haar-like feature consists of two or three jointed “black” and “white” rectangles:

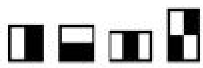

Figure 1: A set of basic Haar-like features.

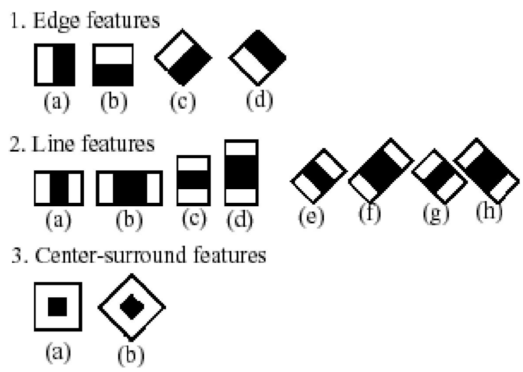

Figure 2: A set of extended Haar-like features.

The value of a Haar-like feature is the difference between the sum of the pixel gray level values within the black and white rectangular regions:
f(x)=Sum_black rectangle (pixel gray level) – Sum_white rectangle (pixel gray level)

Compared with raw pixel values, Haar-like features can reduce/increase the in-class/ out-of-class variability, and thus making classification easier.

Scan image at multiple positions and scales as in previous approaches. Apply Adaboost strong classifier (which is based on Rectangle Features) to decide whether the search window contains a face or not. Adaboost Strong Classifier: linear combination of weak classifiers. H(x) = 1 if window x has a face and 0 otherwise.

$$H(x) = \begin{cases} 1 \text{ if } \sum_{t=1}^{T} \alpha_t h_t(x) \geq \phi \\ 0 \text{ otherwise} \end{cases}$$

Threshold
Weak Classifier
Weights

$$h_j(x) = \begin{cases} 1 \text{ if } p_j f_j(x) < p_j \theta_j \\ 0 \text{ otherwise} \end{cases}$$

Threshold
Sign
Rectangle Feature

Each week classifier corresponds to a single Rectangle Feature. Adaboost ensembles many weak classifiers into one single strong classifier
1. Initialize sample weights
2. For each cycle: Find a classifier/rectangle feature that performs well on the weighted samples. Increase weights of misclassified examples.
3. Return a weighted combination of classifiers
Attentional Cascade
1. We start with simple classifiers which reject many of the negative subwindows
2. While detecting almost all positive sub-windows
3. Positive results from the first classifier triggers the evaluation of a second (more complex) classifier, and so on
4. A negative outcome at any point leads to the immediate rejection of the subwindow
Each frame will be processed by haar classifier and then each profile face will be found. For this project, we use haar-like to train a classifier to detect face area.

4. SVM Classification
An SVM model is a representation of the examples as points in space, mapped so that the examples of the separate categories are divided by a clear gap that is as wide as possible. New examples are then mapped into that same space and predicted to belong to a category based on which side of the gap they fall on.

In this project, we build up three binary svm models to classify four different facial expressions. When a test frame enters, the first model will classify it into happy emotion or other emotions. If the frame falls into happy, the classification is finished and facial expression recognition result is obtained. If not, the frame will continue to be tested by the second model. As a result, if it belongs to sad emotion, we complete the classification and the corresponding facial expression is sadness. Or we need to continue our classification and we can get the result after the frame is processed by the third model.

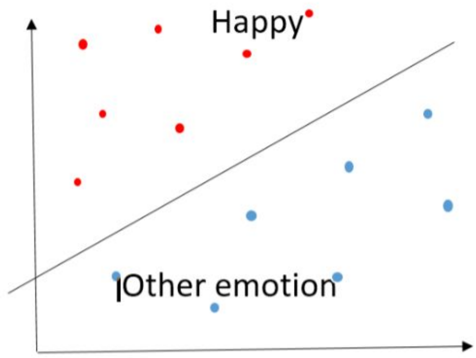

Figure 3: Frame processed by first model

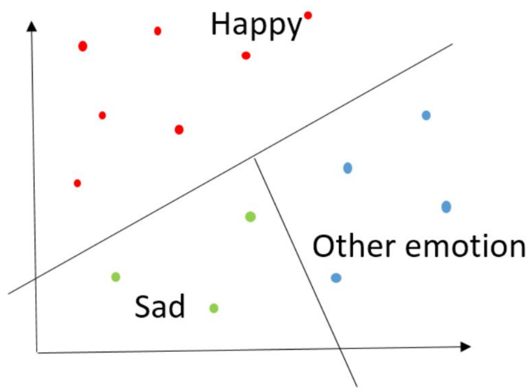

Figure 4: Frame processed by second model

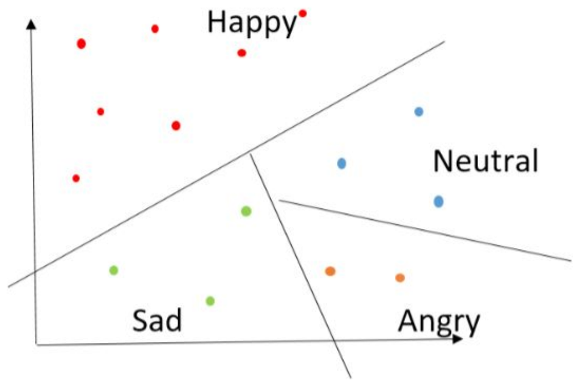

Figure 5: Frame processed by third model

5. Results & Conclusions

Facial Expression	Accuracy
Happy	0.983051
Sad	0.985294
Angry	0.947368
Neutral	0.72
Total	0.918803

Table 2 Result of Accuracy

It was observed that result was directly proportional to intensity of training provided. It was also observed that result was robust to pose variations; light intensity changes and is person independent model. Also the result was robust to background changes.

Happy	Sad	Angry	Neutral	I/O
58	0	0	7	Happy
1	67	3	7	Sad
0	1	54	0	Angry
0	0	0	36	Neutral

TABLE 3 Confusion Matrix for the Facial Expression Classifier as Measured on our Image Set

ang	dis	fea	hap	sad	sur	I/O
88	13	13	6	6	0	ang
0	73	0	6	6	0	dis
0	0	73	6	0	8	fea
0	0	0	78	0	0	hap
13	13	0	0	81	0	sad
0	0	0	0	0	85	sur
0	0	13	6	6	8	neu

TABLE 4 Compared Confusion Matrix for the Facial Expression Classifier as Measured on Ekman Image Set

Compared to table 3, Table 4 shows a confusion matrix showing misclassification rates for expressions based only on the Japanese female models.

6. References

[1] L. S. Chen. Joint processing of audio-visual information for the recognition of emotional expressions in human-computer interaction. PhD thesis, University of Illinois at Urbana-Champaign, Dept. of Electrical Engineering, 2000.

[2] L. S. Chen, H. Tao, T. S. Huang, T. Miyasato, and R. Nakatsu. Emotion recognition from audiovisual information. In Proc. IEEE Workshop on Multimedia Signal Processing, pages 83–88, Los Angeles, CA, USA, Dec. 7-9, 1998.

[3] L. C. De Silva, T. Miyasato, and R. Natatsu. Facial emotion recognition using multimodal information. In Proc. IEEE Int. Conf. on Information, Communications and Signal Processing (ICICS'97), pages 397– 401, Singapore, Sept. 1997.

[4] P. Ekman and W. V. Friesen. Facial Action Coding System: Investigator's Guide. Consulting Psychologists Press, Palo Alto, CA, 1978.

[5] K. Mase. Recognition of facial expression from optical flow. IEICE Transactions, E74(10):3474–3483, October 1991.

[6] T. Otsuka and J. Ohya. Recognizing multiple persons' facial expressions using HMM based on automatic extraction of significant frames from image sequences. In Proc. Int. Conf. on Image Processing (ICIP-97), pages 546–549, Santa Barbara, CA, USA, Oct. 26-29, 1997.

[7] M. Rosenblum, Y. Yacoob, and L.S. Davis. Human expression recognition from motion using a radial basis function network architecture. IEEE Transactions on Neural Network, 7(5):1121–1138, September 1996.

[8] Y. Yacoob and L.S. Davis. Recognizing human facial expressions from long image sequences using optical flow. IEEE Transactions on Pattern Analysis and Machine Intelligence, 18(6):636–642, June 1996.