Fast Facial Expression Recognition using Multiple Pyramid Strategy

Chengjun Yuan cy3yb@virginia.edu

Yijun He yh9vm@virginia.edu

Abstract

Facial expression recognition is both an interesting and important problem. It has many potential applications in data-driven animation, intelligent robotics, clinic psychology, and behavior monitoring. Here, we present a multiple layers pyramid machine learning strategy. The most distinguishing facial expression: happiness and sadness are classified using small part of facial features, and then surprise and fear are detected by involving large part of facial features, and at last, anger and disgust are recognized by using all features. This kind of multi-layer structure is expected to reduce the computational complexity of SVM with little decrease in accuracy. The famous facial expression dataset - Cohn and Kanade AU-coded expression database will be utilized to explore how our strategy performs in reaching a performance balance between efficiency and accuracy.

1 Introduction

The face is the fundamental part of day to day interpersonal communication. Humans use the face along with facial expressions to express their emotional states (anger, surprise, happiness, etc.) to accompany and enhance the meaning of their thoughts. The importance of automatically recognizing facial expressions by using machine learning technique is apparent and can be beneficial to many different scientific subjects such as clinic psychology, neurology, as well as, applications of everyday life such as driver monitoring systems and automated tutoring systems[1, 2].

One essential step for successful facial expression recognition is to construct an effective facial representation from face images. The most common used model for facial representation is facial action coding system (FACS)[3]. It introduced Action Units (AU) to express facial movements. Action Units are a set of actions that correspond either to muscle movement in facial expressions such as raising upper lip or blinking, or some miscellaneous actions such as bite lip or blow. The FACS consists of up to 46 action units. However, as we know, more information of facial expression is appeared in the eyes, eyebrows and mouth region than other regions. These regions are called "salient regions" where "salient" means most noticeable or most important.

In addition, a psycho-visual experimental study[4] suggests that for happiness, sadness and surprise expression, only one facial region is salient, while for other expressions (anger, fear, disgust) two facial region are salient. For example, mouth is the most important region for representing surprise expression, but anger facial expression is difficult to determine if only from mouth region. A pyramid strategy based on this psycho-visual study has been proposed and tested to achieve improvement in accuracy and velocity of recognition[5]. However, the pyramid has only two layers. In the first layer, it is classified as happiness, sadness or surprise by the features extracted from the mouth region. If none of them fit the classifier, then the other three expressions: anger, fear and disgust are classified by the features of both mouth region and eyes region. If two layers pyramid strategy can bring benefits, how about three or more layers pyramid? No answers to this question has been given according to our best literature search. So here we will examine the multiple pyramid strategy in acceleration of the high-accuracy facial expression recognition.

2 Previous work on FER

In general, there are three steps for FER: detection of face region in the image, extraction of facial information as features, and classification of the facial features as a particular facial expression such as anger, happiness or fear et al.[6]. Here, we only focus on the third step. A wide range of techniques have been employed in the classification of facial expression, including linear discriminant analysis (LDA)[7], nearest neighbor classification[8], support vector machine (SVM), Bayes classifier[9], neural network[10] as well as Adaboost classification[11] with a selection of weak classifiers. However, there are also weakness in these classifiers. The neural network needs many adjustable parameters and the learning time is long. The SVM's operational efficiency is low and it needs huge computational resources[12]. The effect of Adaboost depends on the week classifier selection. Two recent research works made progress in reducing the computational cost for FER. One is to use Adaboost for features selection and then followed by SVM classifier[13]. Since less features are input into SVM, the recognition efficiency is improved with high accuracy still remained. The other is to adopt Pyramid local phase quantization descriptor (PQD). Since only partial facial information (eyes and mouth region) are extracted for recognition and two layer pyramid strategy is implemented, it provided a fast recognition with the accuracy up to 96.7%[5].

3 Why is It Related to Machine Learning?

Here we will build an automated model for FER based on SVM technique. The model will be trained in training dataset that gives AU features and the corresponding expressions, and tested in test dataset. 10-fold cross validation method will be adopted to evaluate its accuracy. We will carefully choose and adjust the three layers pyramid strategy to find the balance point between the recognition efficiency and accuracy of our model.

4 Method & Experimental Design

4.1 Dataset

The dataset for our training and test is the extended Cohn and Kanade AU-coded expression database[14, 15]. It includes 592 sequences from 123 posers. Each sequence begins with a neutral expression and proceeds to a peak expression. The peak expression for each sequence is fully FACS coded with 43 AUs and given an emotion label among 7 emotion expressions as detailed shown in TABLE 1 and TABLE 2.

TABLE 1. The AUs coded by FACS on the CK+ database.

AU	Name	AU	Name	AU	Name
1	Inner Brow Raiser	14	Dimpler	27	Mouth Stretch
2	Outer Brow Raiser	15	Lip Corner Depressor	28	Lip Suck
4	Brow Lowerer	16	Lower Lip Depressor	29	Jaw Thrust
5	Upper Lid Raiser	17	Chin Raiser	31	Jaw Clencher
6	Cheek Raiser	18	Lip Puckerer	34	Cheek Puff
7	Lid Tightener	20	Lip stretcher	38	Nostril Dilator
9	Nose Wrinkler	22	Lip Funneler	39	Nostril Compressor
10	Upper Lip Raiser	23	Lip Tightener	41	Lid droop
11	Nasolabial Deepener	24	Lip Pressor	42	Slit
12	Lip Corner Puller	25	Lips part	43	Eyes Closed
13	Cheek Puffer	26	Jaw Drop		

TABLE 2. Frequency of emotions.

No.	Emotion	N
1	Angry	45
2	Contempt	18
3	Disgust	59
4	Fear	25
5	Happiness	69
6	Sadness	28
7	Surprise	83

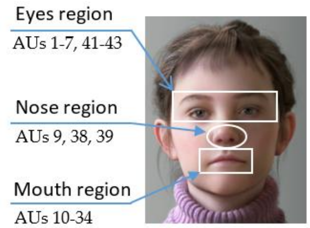


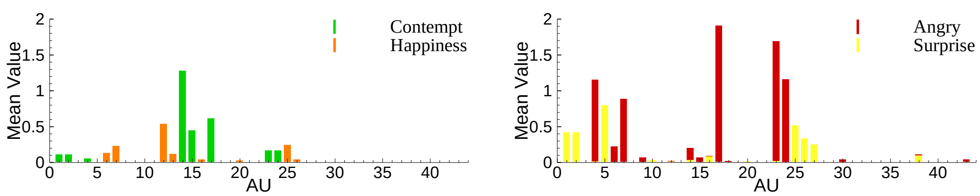
Figure 1. Three regions that AUs are distributed.

As mentioned before, most facial features are focused on the eyes region and mouth region. Among the 32 used AUs, there are 9 items related with the eyes region and 20 items in the mouth region. The rest 3 AUs are located at the nose region. It is worth noting that several AUs include features of both the nose and mouth regions, such as AUs 10 – 15 that are classified here to the mouth region. The whole dataset can be freely downloaded from the website http://www.pitt.edu/~emotion/ck-spread.htm.

4.2 Three layers pyramid ML strategy (Proposed Method)

The Three layers pyramid strategy for facial expression recognition in this paper intends to reduce the feature vector dimensionality and so bring in a decrease in computational complexity. Detailed steps are listed below and are illustrated in Figure 3:

Before classification: There are three groups of emotion. The first group includes contempt and happiness. The second group contains surprise and angry, and the last group have fear, sadness and disgust. This strategy comes from the analysis of section 4.3. According to TABLE *3*, happiness and contempt are mainly expressed by mouth region. By comparison of their distributions of AUs shown in Figure 2 (a), they have different set of AU features, which is beneficial for classification of them. For the second level, there is also this kind of discipline as shown in Figure 2 (b). They are the additional reasons we set the first two levels of emotions besides the psycho-visual study mentioned before.



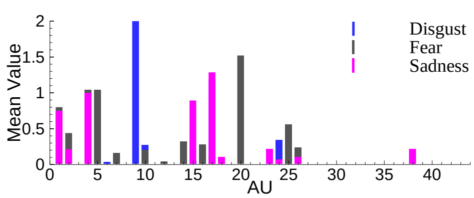


Figure 2. Integrated distributions of UA features of each level of emotions. (a) The first level. (b) The second level. (c) The third level.

Step 1: The AU features from the mouth region are used to classify whether the emotion belongs to first group or other two groups. If yes, then it will be classified either as contempt or happiness. If not, then go to step 2.

Step 2: Here features from both mouth and eyes regions are fed into SVM classifier to judge whether it is in group 2 or not. If yes, then it may be either surprise or angry expression. If no, then go to step 3.

Step 3: All features from the whole face are utilized in the classifier whether it is fear, sadness or disgust.

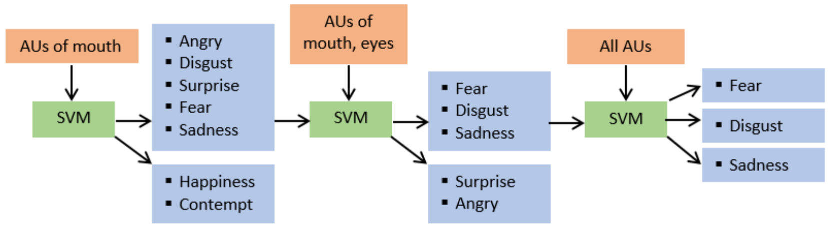


Figure 3. The framework of 3 layers pyramid machine learning strategy.

In this work, we adopt SVM as the classifier for expression recognition with seven basic emotions. SVM performs an implicit mapping of data into a higher dimensional feature space, and then finds a linear separating hyper-plane with the maximal margin to separate data in this higher dimensional space. Given a training set (x_i, y_i), $\mathrm{i} = 1, \dots, \mathrm{l}$, the new test example x is classified as:

$$f(x) = sgn\left(\sum_{i=1}^{l} \alpha_i y_i K(x_i, x) + b\right) \quad (1)$$

Where α_i are Lagrange multipliers of dual optimization problems that describe the separating hyper-plane, $K(x_i, x)$ is the kernel function and b is the threshold parameter of hyper-plane. The most frequently used kernel functions are the linear, polynomial, and Radial basis function (RBF) kernels:

$$Linear\ Kernel: K(x, y) = x \cdot y \quad (2)$$

$$Polynomial\ \ Kernel: K(x, y) = (x \cdot y)^d \quad (3)$$

$$RBF\ \ Kernel: K(x, y) = exp\left(-\frac{\|x - y\|^2}{2\sigma^2}\right) \quad (4)$$

These three kernels will be evaluated in section 4.3 for model selection design.

4.3 Model selection design

Each emotion has its specific facial features represented by AUs. In order to extract the corresponding features, here the mean values of 43 AUs of all sequences of each emotion are calculated by the equation

(5) below. Their distributions are plotted in Figure 4. It can be seen clearly that the major features of angry are AUs 4, 7, 17, 23 and 24, which means all major AUs are located in the eyes and mouth regions.

$$\overline{AU_i} = \frac{1}{N}\sum_{j}^{N} AU_{i,j} \quad (5)$$

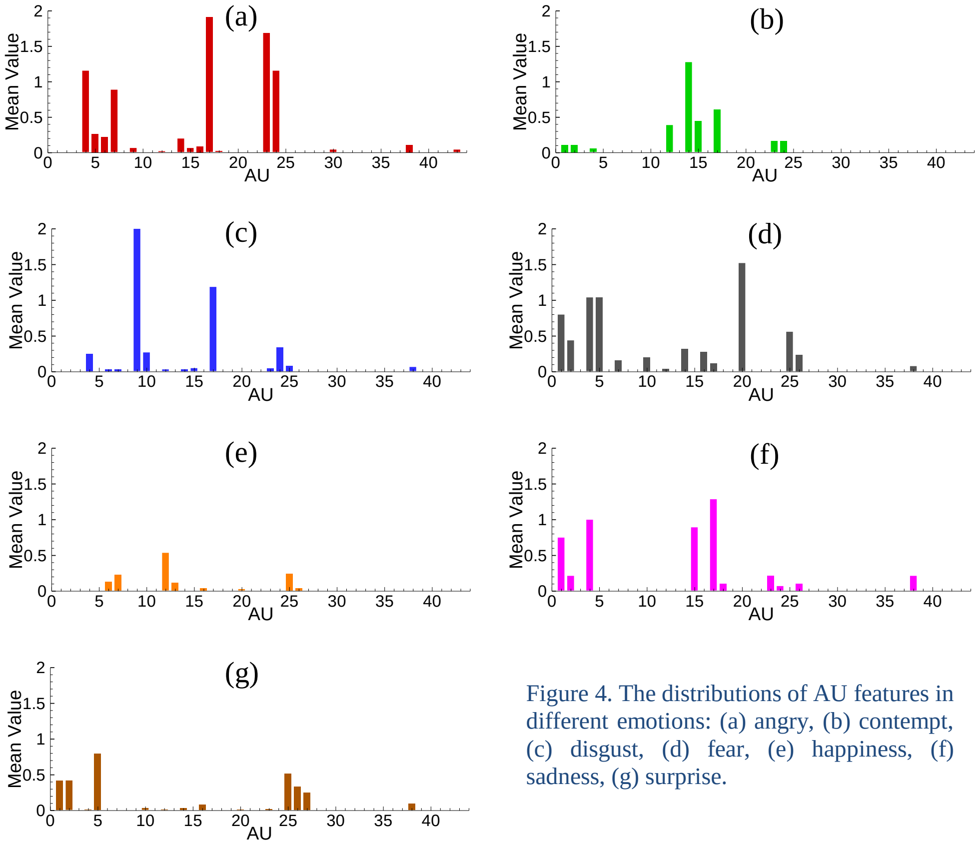


Figure 4. The distributions of AU features in different emotions: (a) angry, (b) contempt, (c) disgust, (d) fear, (e) happiness, (f) sadness, (g) surprise.

The similar analysis has been done for the other emotions, and the results are saved in TABLE 3. The contempt and happy emotions are mainly presented in the mouth region. The angry, fear, sadness and surprise emotions are featured by the eyes and mouth regions together. Only the disgust emotion is primarily located in the nose and mouth regions.

TABLE 3. Emotion description in terms of major AUs and regions

Emotion	Major AUs	Major Regions
Angry	4,7,17,23,24	Eyes, Mouth
Contempt	12,14,15,17	Mouth
Disgust	9,17,24	Nose, Mouth
Fear	1,2,4,5,20,25	Eyes, Mouth
Happiness	12	Mouth
Sadness	1,4,15,17	Eyes, Mouth
Surprise	1,2,5,25	Eyes, Mouth

5 Our Team Members

Chengjun Yuan has hands-on experiences on numerical algorithm optimization and data analysis, including two peer-reviewed publications and several machine-learning related projects, such as the spam email filter system by combining K-means Cluster and SVM, and Mining Code Blocks-functionalities Map using structured information retrieval & SVM (ongoing). Considering his solid skills in algorithm design and machine learning techniques, he is very eligible to make reasonable progress in the topic of this paper.

Yijun He is taking machine learning course this semester, and he has learnt much about statistic learning tools such as LDA, SVM. He is very interested in application of machine learning techniques on facial expression recognition and looking forward to do his best in this project.

References

1. Koutlas, A. and D.I. Fotiadis, *Image Processing and Machine Learning Techniques for Facial Expression Recognition.* IGI Global (January 2009), 2009: p. 1-16.
2. Zhao, W., et al., *Face recognition: A literature survey.* ACM computing surveys (CSUR), 2003. **35**(4): p. 399-458.
3. Ekman, P. and W. Friesen, *Thefacial action coding system: A technique for the measurement offacial movement*. 1978, Palo Alto, CA: Consulting Psychologists Press.
4. Khan, R.A., et al., *Framework for reliable, real-time facial expression recognition for low resolution images.* Pattern Recognition Letters, 2013. **34**(10): p. 1159-1168.
5. Vo, A. and N.Q. Ly, *Facial Expression Recognition Using Pyramid Local Phase Quantization Descriptor*, in *Knowledge and Systems Engineering*. 2015, Springer. p. 105-115.
6. Sandbach, G., et al., *Static and dynamic 3D facial expression recognition: A comprehensive survey.* Image and Vision Computing, 2012. **30**(10): p. 683-697.
7. Belhumeur, P.N., J.P. Hespanha, and D.J. Kriegman, *Eigenfaces vs. fisherfaces: Recognition using class specific linear projection.* Pattern Analysis and Machine Intelligence, IEEE Transactions on, 1997. **19**(7): p. 711-720.
8. Lozano-Monasor, E., et al., *Facial expression recognition from webcam based on active shape models and support vector machines*, in *Ambient Assisted Living and Daily Activities*. 2014, Springer. p. 147-154.
9. Savran, A., B. Sankur, and M. Bilge, *Comparative evaluation of 3D versus 2D modality for automatic detection of facial action units.* Pattern Recognit. v45 i2, 2011: p. 767-782.
10. Takahashi, K., et al., *Remarks on Computational Facial Expression Recognition from HOG Features Using Quaternion Multi-layer Neural Network*, in *Engineering Applications of Neural Networks*. 2014, Springer. p. 15-24.
11. An, K.H. and M.J. Chung. *Learning discriminative MspLBP features based on Ada-LDA for multi-class pattern classification*. in *Robotics and Automation (ICRA), 2010 IEEE International Conference on*. 2010. IEEE.
12. Liu, Z.-T., et al. *A novel facial expression recognition method based on extreme learning machine*. in *Control Conference (CCC), 2015 34th Chinese*. 2015. IEEE.
13. Bartlett, M.S., et al. *Recognizing facial expression: machine learning and application to spontaneous behavior*. in *Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on*. 2005. IEEE.
14. Kanade, T., J.F. Cohn, and Y. Tian. *Comprehensive database for facial expression analysis*. in *Automatic Face and Gesture Recognition, 2000. Proceedings. Fourth IEEE International Conference on*. 2000. IEEE.
15. Lucey, P., et al. *The Extended Cohn-Kanade Dataset (CK+): A complete dataset for action unit and emotion-specified expression*. in *Computer Vision and Pattern Recognition Workshops (CVPRW), 2010 IEEE Computer Society Conference on*. 2010. IEEE.