A Machine Learning Algorithm of Human-Computer-Interface Application

-An AdaRank Model Approach to Facial Expression Recognition

Chang-Yi Kao
Institute for Information Industry
Taipei, Taiwan 10607, Republic of China
E-mail: edenkao@iii.org.tw

Chin-Shyurng Fahn
National Taiwan University of Science and Technology
Taipei, Taiwan 10607, Republic of China
E-mail: csfahn @mail.ntust.edu.tw

***Abstract*—A completely automatic facial expression recognition system is presented in this paper, which consists of three main procedures. The first is based on skin color blocks and geometrical properties applied to eliminate the skin color regions that do not belong to the face in the HSV color space. Than we find proper ranges of eyes, mouth, and eyebrows according to the positions of pupils and center of a mouth. Subsequently, we perform both the edge detection and binarization operations on the above ranged images to obtain 16 landmarks. After manipulating these landmarks, 16 characteristic distances are the facial feature produced to represent a kind of expressions. Finally, we subtract the 16 characteristic distances of a neutral face from the 16 characteristic distances of a certain expression to acquire its 16 displacement values fed to a classifier with an incremental learning scheme, which can identify six kinds of expressions: joy, anger, surprise, fear, sadness, and neutral. We choose the AdaRank model as the core technique to implement our strong facial expression classifier. Our model, referred to as AdaRank, repeatedly constructs classifiers on the basis of re-weighted training data and finally linearly combines the classifiers for making ranking predictions. Through conducting many experiments, the statistics of performance reveals that the accuracy rate of our facial expression recognition system reaches more than 95%.**

Keywords- facial expression recognition, landmarks, facial feature, AdaRank

I. INTRODUCTION

Recently, more effective and friendly methods for HCI have been developed; such methods, unlike the conventional ones, do not rely on traditional input devices; for instance, keyboards and mice. We want to realize this dream and the facial expression recognition seems to be indispensable technology. The facial expression recognition will be a very important technique in the future life. Nowadays, more and more digital products employ the technology of facial expression recognition.

The completed face tracking and recognition researches could be divided into three general procedures briefly: face tracking [1], and face recognition [2]. There are many problems which are closely related to face detection. Face localization is a simplified problem of face detection on the assumption that there is only one face in an input image. Face recognition is to compare an input face against a library of known faces and reports the result if a match is found. More and more researches have been devoted to the face detection and facial expression recognition of an image sequence [3-5]. The task of human face image analysis includes face detection, face recognition, and facial expression recognition. Colmenarez et al. [6] presented a Bayesian probabilistic approach to recognizing faces and facial expressions. They possessed the mutual benefits in similarity measures between faces and facial expressions. Nevertheless, since each person in a database must have a trained model, this limits its realistic use in facial expression analysis. Qin and He [7] took the technological advantages of both the support vector machine (SVM) and Gabor feature extraction for face recognition. First, the SVM is successfully applied to face recognition using the Gabor features of key points; second, such Gabor features are introduced to represent a whole face in a computable dimensional space. Given a set of images, the aforementioned key points are manually labeled with the same positioning principle on target images. The system proposed in [8] automatically detected frontal faces in an image sequence and classified them into seven classes in real time: neutral, anger, disgust, fear, joy, sadness, and surprise. Its expression recognizer receives image regions produced by a face detector and then a Gabor representation of the facial image region is formed to be later processed by a bank of SVM-based classifiers. A facial expression feature stream was used to train a parallel HMM structure in a similar fashion explained in [9], which provided a probabilistic model for temporal recurrent facial expression patterns. The segments corresponding to these patterns are detected and labeled over the training image sequences, where labels are outputted as dynamic facial expression patterns.

There are several topics in this field which are related to this work. They are boosting, learning to rank, and direct optimization of performance measures. Typical methods of the approach include Ranking SVM [10], RankBoost [11], RankNet [12], and some other approaches to learning to rank [3]. AdaRank is also one that tries to directly optimize multivariate performance measures. AdaRank is unique in that it employs an exponential loss function boosting technique.

The architecture and detail of our facial expression recognition system described in next sections, and the method that we proposed is organized as follows:

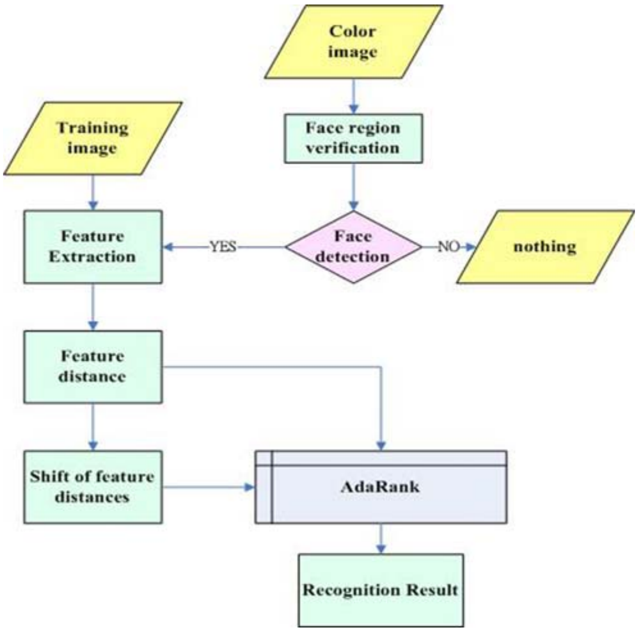

Figure 1. The facial expression recognition system architecture.

II. FACE DETECTION AND FEATURE EXTRACTION

A. Face Detection

In order to enable the system to extract facial features more efficiently and more perfectly, we must narrow the range of face detection first. In our system design philosophy, the skin color cue is an obvious characteristic to detect human faces. In this phase, we will elaborate how to look for possible face candidate regions in an image, and further search the rough positions of eyes and mouth, used for facial expression recognition later. To accomplish this, we will integrate skin color detection, morphological dilation operation, and facial feature detection. Subsequently, a filtering operation based on geometrical properties is applied to eliminate the skin color regions that do not belong to the face; for example, arms and hands. When system is started, the camera captures images immediately, and then lodged in the system, which can deal with every frame. Finally, most possible area of a human face is extracted from an images sequence.

B. Facial Feature Extraction

First, we label the low boundary of a face region by the following equation.

$$F_L = \begin{cases} F_L, & \text{if } B_H / B_W \le 1.4 \\ F_U + 1.4 * B_W, & \text{if } B_H / B_W > 1.4 \end{cases} \quad (1)$$

Where FL and FU are the low and up boundaries of a face region, individually.We draw proper ranges of eyes, mouth, and eyebrows according to the positions of pupils and center of a mouth. Then we perform both the edge detection and binarization operations on the above ranged images to obtain the characteristic information of facial expressions. Finally, we acquire the feature points from the characteristic information. We use the 16 feature points (landmarks) used in our facial expression recognition system.

Landmarks of Eyes Extraction: We first draw proper ranges of eyes using the positions of pupils and then obtain their edge detection and binarization images. At last, we extract eight landmarks of the both eyes.

Landmarks of Eyebrows Extraction: It is a little difficult to extract landmarks of eyebrows, because eyebrows often shift their positions along with distinct expressions such as happiness, anger, sadness, and joy. On the other hand the appearance of an eyebrow has a lot of changes. Hence the position and range of an eyebrow are hard to define in any situation. We utilize the color difference of eyebrows and the skin. Because eyebrows are darker than the skin we can adopt this property to extract the landmarks of eyebrows. We first draw the proper ranges of eyebrows using the positions of pupils and then obtain their edge detection and binarization images. At last, we extract four landmarks of the two eyebrows.

Landmarks of Mouth Extraction: According to our observation and experiments, the lip color is usually darker than the skin color, so we apply this characteristic to extract the landmarks of a mouth. To achieve this, we first draw a proper range referring to the central point of a mouth and then obtain the edge detection and binarization images. At last, we extract four landmarks of the mouth.

III. FACIAL EXPRESSION RECOGNITION

We utilize the landmarks extracted previously to conduct facial expression recognition. The main technique of classification is based on the difference between various kinds of facial expressions and a neutral facial expression. From the 16 landmarks of a human face, we produce 16 characteristic distances which represent a kind of expressions. Then we subtract the 16 characteristic distances of a neutral face from the 16 characteristic distances of a certain expression to acquire its 16 displacement values. Finally, we employ AdaRank techniques to recognize five kinds of expressions.

A. Feature Distances

According to our observation, the landmarks of eyebrows and mouths will produce more obvious displacement for the facial expressions excluding the neutral. For example, when the people are joying, the facial landmarks on the left and right bounds of a mouth are raised up and drawn apart to both sides, while the people are angry, the facial landmarks on the inside bounds of eyebrows are pressed inwards and downwards. Because the face areas affected by each kind of facial expressions are not the same, in order to recognize facial expressions effectively, we must understand the corresponding relation between the facial expressions and displacement of the face areas. We definition the distinguishing features of different facial expressions detail descriptions for five kind of facial expression:

Joy. (1). the corners of a mouth are raised up. (2). the width of a mouth becomes large. (3). eyes are a little diminished.

Anger. (1). two eyebrows are close to each other. (2). the interval between eyebrows appears vertical lines. (3). eyes open widely.
Surprise. (1). the eyebrows are raised up. (2). the chin is fallen down. (3). the height of a mouth becomes large.
Fear. (1). two eyebrows are close to each other or raised up. (2). the width of a mouth becomes large. (3). eyes open widely.
Sadness. (1). the corners of a mouth are fallen down. (2). two eyebrows are a little close to each other. (3). eyes are a little diminished. (4). the upper lip is carried up. Next, with reference to the 16 landmarks on a human face, we produce 16 characteristic distances as Figure 2. shows. These values of 16 characteristic distances are the main features used for recognizing facial expressions.

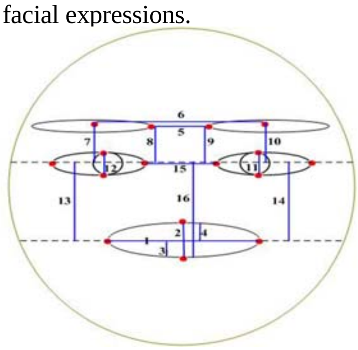

Figure 2. 16 characteristic distances derived from the 16 landmarks on a human face.

B. *AdaRank Algorithm*

In this model, this ranking mathematical model was constructed based on Metric properties. AdaRank is a simple yet powerful method. More importantly, it is a method that can be justified from the theoretical viewpoint. The following description was the algorithm [14]. AdaRank takes a training set S as input and takes the performance measure function E and the number of iterations T as parameters. The input is Eq. (2).

$$S = \{(q_i, d_i, y_i)\}_{i=1}^{m} \tag{2}$$

It will update the Eq. (3) iteratively from *i* to *t*.

$$p_{t=1}(i) = \frac{\exp\{-E(\pi(q_i, d_i, f_i), y_i)\}}{\sum_{j=1}^{m} \exp\{-E(\pi(q_i, d_i, f_i), y_i)\}} \tag{3}$$

Training error will be continuously reduced during learning phase. The following bound holds on the ranking accuracy of the AdaRank algorithm on training data:

$$\begin{aligned} &\frac{1}{m}\sum_{j=1}^{m} \exp\{-E(\pi(q_i, d_i, f_i), y_i)\} \\ &\geq 1 - \prod_{t=1}^{T} e^{-\delta^t} \min \sqrt{1-\varphi(t)^2} \end{aligned} \tag{4}$$

Where, each condition constrains as Eq. (5), (6), (7).

$$\varphi(t) = \sum_{i=1}^{m} P_t(i) E(\pi(q_i, d_i, h_t), y_i) \tag{5}$$

$$\delta_{\min}^{t} = \min_{i=1,\dots,m} \delta^{t} \tag{6}$$

$$\begin{aligned} \delta_i^t &= E(\pi(q_i, d_i, f_{t-1} + \alpha_t h_t), y_i) - \\ &E(\pi(q_i, d_i, f_{t-1t}), y_i) - \\ &\alpha_t E(\pi(q_i, d_i, h_t), y_i) \end{aligned} \tag{7}$$

For all $i = 1, 2,\dots, m$ and $t = 1, 2,\dots, T$.

In the implementation, as weak classier we choose the feature that has the optimal weighted performance among all of the features:

$$\max_{k} \sum_{i=1}^{m} P_i E(\pi(q_i, d_i, x_i), y_i) \tag{8}$$

We use AdaRank algorithm for training the initialized parameters of learning to rank to provide the trained rank. After training procedure, we use the trained rank parameters and discrete vector, hand motion trajectory, as input to obtain the best path. By the best path and facial expression recognition database, we can recognize the facial expression path.

Comparatively, the AdaRank algorithm has an advantage of facilitating the speed of convergence. Thus, we can update the training samples to deal with comprehensive circumstances of using varied expression features, but need not spend much computational cost. We choose the AdaRank algorithm as the core technique to implement our strong facial expression classifier. Moreover, the algorithm support the recognition system a bottom-up ranking classification structure for the expression recognition, even we do not implement it in this research at this paper now.

IV. EXPERIMENTAL RESULT

We will compare three different classifiers of recognizing facial expressions using MLPs, SVMs, and ABAs. Finally, we will show the sequential composite facial expressions recognition result.

A. *The Facial Expression Database*

We set up one small-scale database of facial expressions by ourselves using the web camera “Logitech Quick-Cam Pro 4000” to take image sequences with the resolution of 320 × 240 pixels. In addition, this database employed in our system is different from the ordinary databases of facial expressions. Such databases usually store static face images one by one.

At present, the facial features of ten persons (eight males and two females) are stored in our database, and each person has 1,200 materials comprising joy, anger, surprise, fear, sadness, and neutral, each of which contains 200 materials. That is, there are 12,000 materials totally in our facial expression database. In our experimental, there are some image samples of six kinds of expressions, where the faces may be panned from -30° to 30° and tilted from -10° to 10°.

B. *The Results of Facial Expression Recognition*

The AdaRank algorithm we adopt is a ranking classifier. In the future, we will also propose a bottom-up hierarchical classification structure for multi-class facial expression recognition using AdaRank algorithm. Such a decision tree of recognizing the six kinds of expressions is similar to that based on SVMs. The total training time for the condition of 12,000 samples by AdaRank algorithm is about 4 minutes. The detailed experimental data are recorded in the follow Table. And we can discover that the accuracy rates of recognizing expressions are better and evener than those resulting from SVMs. The accuracy rate of recognizing expression "surprise" is also the highest. But, the accuracy rate of recognizing expression "fear" is lower. It is caused by the landmarks of a mouth in the expression "fear" similar to those in the expression "joy." On the other hand, we can also find that the accuracy rate of recognizing expression "sadness" has a better outcome.

TABLE I. THE FACIAL EXPRESSION RECOGNITION RESULTS OF ADARANK ALGORITHMS

Recognition result / Expression type	Neutral	Joy	Anger	Surprise	Fear	Sadness	Other
Neutral	1918	2	22	3	15	28	7
Joy	24	1931	0	1	30	6	1
Anger	7	2	1978	0	17	6	3
Surprise	0	0	0	1993	5	0	0
Fear	5	118	6	0	1837	15	19
Sadness	26	8	39	0	9	1918	16
Other	2	4	4	2	8	11	1981

The precision rate is slightly different from the rec all rate. In the denominator of these two rates, besides the true positive, the precision rate includes the false positive but the recall rate has the false negative. For the precision rate, a higher precision rate represents that the number of false positives is few. As to the recall rate, the smaller the number of false negatives is, the higher recall rate obtains. It means that the most expressions can be classified correctly.

We then apply these two rates, recall rate and precision rate, to measure the performance of our facial expression recognition system. Table 2 records the precision and recall rates of the experiments mentioned earlier. And the Average accuracy of these three method show as Table 2.

TABLE II. THE AVERAGE ACCURACY RATES OF THREE CLASSIFIERS

	Multi-Layer Perceptrons	Support Vector Machines	AdaRank Algorithms
Average accuracy	95.7 %	92.2 %	96.8 %

V. CONCLUSIONS

We can observe that the accuracy rate of recognizing expression "surprise" is higher than those of recognizing the other five kinds of expressions using the three classifiers, because when people are surprised, the facial landmarks are a lot of change; for instance, the eyebrows are raised up, the chin is fallen down, and the height of a mouth becomes large. However, the accuracy rates of recognizing expressions "fear" and "sadness" are lower, because the facial landmarks of expression "sadness" are not obviously shifted and the landmarks of a mouth in the expression "fear" are similar to those in the expression "joy." Thus, recognizing the two kinds of expressions yields lower accuracy rates.On the whole, the classification result from the Gentle AdaRank algorithms is better than that from the support vector machines. And our future works are worth investigating to attain better performance. Because the performance of AdaRank algorithm is a little slightly worse than other boosting algorithm.

ACKNOWLEDGMENT

This study is conducted under the "Project Digital Convergence Service Open Platform" of the Institute for Information Industry which is subsidized by the Ministry of Economy Affairs of the Republic of China.

REFERENCES

[1] K. C. Fan, Y. K. Wang, and B. F. Chen, "Introduction of tracking algorithms," Image and Recognition, Vol. 8, No. 4, pp. 17-30, 2002.

[2] W. Zhao, R. Chellappa, P. J. Phillips, and A. Rosenfeld, "Face recognition: a literature survey," ACM Computing Surveys, vol. 35, no. 4, pp. 399-458, 2003.

[3] Yang, M.H., Kriegman, D., and Ahuja, N.: 'Detecting faces in images: a survey', IEEE Trans. Pattern Anal. Machine Intell., 2002, 24, (1), pp. 34-58

[4] Yang, G., and Huang, T.S.: 'Human face detection in a complex background', Pattern Recogni., 1994, 27, (1), pp. 53-64

[5] Yow, K.C., and Cipolla, R.: 'Feature-based human face detection', Image Vision Comput., 1997, 15, (9), pp. 712-735

[6] S. L. Phung, A. Bouzerdoum, and D. Chai, "A novel skin color model in YCbCr color space and its application to human face detection," in Proceedings of IEEE International Conference on Image Processing, vol. 1, pp. 289-291, 2002.

[7] C. Lin, "Face detection, pose classification, and face recognition based on triangle geometry and color features," Master Paper, Department of Computer Science Information Engineering, National Central University, Jhongli, Taiwan, 2001.

[8] M. Soriano, S. Huovinen, B. Martinkauppi, and M. Laaksonen, "Using the skin locus to cope with changing illumination conditions in color-based face tracking," in Proceedings of the IEEE Nordic Signal Processing Symposium, pp. 383-386, 2000.

[9] Sargin, M.E., Erzin, E., Yemez, Y., Tekalp, A.M., Erdem, A.T., Erdem, C., and Ozkan, M.: 'Prosody-driven head-gesture animation', Proc. Int. Conf. Acoust., Speech, and Signal Process., 2007, vol. 2, pp. 677-680

[10] R. Herbrich, T. Graepel, and K. Obermayer. Large Margin rank boundaries for ordinal regression. MIT Press, Cambridge, MA, 2000.

[11] Y. Freund, R. D. Iyer, R. E. Schapire, and Y. Singer. An efficient boosting algorithm for combining preferences. Journal of Machine Learning Research, 4:933–969, 2003.

[12] C. Burges, T. Shaked, E. Renshaw, A. Lazier, M. Deeds, N. Hamilton, and G. ullender. Learning to rank using gradient descent. In ICML 22, pages 89–96, 2005.

[13] G. Fung, R. Rosales, and B. Krishnapuram. Learning rankings via convex hull separation. In Advances in Neural Information Processing Systems 18, pages 395–402. MIT Press, Cambridge, MA, 2006.

[14] J. Xu and H. Li, AdaRank: A boosting algorithm for information retrieval. SIGIR 2007.

[1]Chang-Yi Kao, a Ph.D. candidate in the Department of Computer Science and Information Engineering at National Taiwan University of Science and Technology.