Real Time Facial Expression Recognition with Adaboost

Yubo WANG, Haizhou AI, Bo WU, Chang HUANG
Dept. of Computer Science and Technology, Tsinghua University,
State Key Laboratory of Intelligent Technology and Systems，Beijing 100084, PR China
Email: ahz@mail.tsinghua.edu.cn

Abstract

In this paper, we propose a novel method for facial expression recognition. The facial expression is extracted from human faces by an expression classifier that is learned from boosting Haar feature based Look-Up-Table type weak classifiers. The expression recognition system consists of three modules, face detection, facial feature landmark extraction and facial expression recognition. The implemented system can automatically recognize seven expressions in real time that include anger, disgust, fear, happiness, neutral, sadness and surprise. Experimental results are reported to show its potential applications in human computer interaction.

1. Introduction

Within the past decade or two, significant effort has occurred in developing methods of facial expression recognition, which is an attractive research issue due to its great potentials in real-life applications, such as human-computer interaction (HCI), emotion computing and digital albums. There are a number of difficulties due to the variation and complexity of facial expression across human population and even the same individual. Sometimes we human beings even make mistakes.

There have been a number of related works on this problem, see [2] for a review. Many of classical pattern recognition methods were used to make expression recognition. For instance, M.J. Lyons et al. [3] adopted PCA and LDA to analyze the expression training sets, and got 92% correctness on JAFFE [1] [14] database; C. Padgett et al. [4] trained a back-propagation neural network, with which the average recognition rate was 86% on Ekman's photos [5]; T. Otsuka et al. [6] used hidden Markov model (HMM) based classifiers to recognize one of six facial expressions near real time. More recently, M.S. Bartlett et al. [7] proposed Gabor feature based AdaSVM method to recognize expression, and obtain a good performance on Cohen-Kanade expression database [8].

In this paper, we propose a novel method for facial expression recognition. The facial expression is extracted from human faces by a multi-class expression classifier that is learned from boosting Haar feature based Look-Up-Table type weak classifiers [11]. The expression recognition system classifies facial expression into seven categories in real time: anger, disgust, fear, happiness, neutral, sadness and surprise.

The rest of this paper is organized as follows: Section 2 presents an overview of our facial expression recognition system; Section 3 introduces the continuous multi-class Adaboost learning algorithm; Section 4 describes the Haar feature based LUT type weak classifiers; Section 5 gives the experiment results of expression recognition; and finally Section 6 gives the conclusion and discussion. The experiments are carried out on the Japanese Female Facial Expression (JAFFE) database [1] [14].

2. System Overview

The proposed facial expression recognition system consists of three modules, face detection, facial feature landmark extraction and facial expression classification, and see Fig.1 for a flowchart. Expressions are divided into seven categories that include anger, disgust, fear, happiness, neutral, sadness and surprise. Here we only consider the problem of frontal facial expression recognition. For face detection, a variation of the cascade detector proposed by Viola and Jones [12] is used. As for facial feature landmark extraction used in face sample normalization, a SDAM [13] method to locate three key points of human face: the pair of eyes and the mouth center is adopted. And finally a classifier for facial expression recognition is learned by boosting Haar feature based Look-Up-Table type week classifiers.

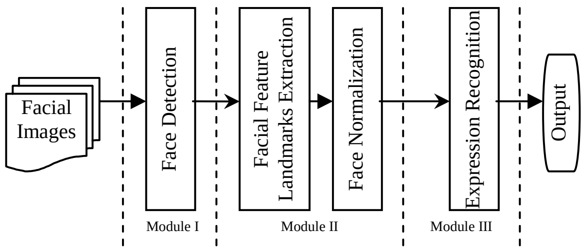


Figure 1: An automatic expression recognition system.

3. The Continuous Multi-class Adaboost Learning Algorithm

Adaboost [9] is a leaning algorithm that selects a small number of weak classifiers from a large weak classifier pool or hypothesis space to construct a strong classifier. The two-class Adaboost algorithm has been successfully used in face detection [10], gender classification [11], etc. However, many problems such as expression recognition are multi-class in nature. Fortunately, based on a measure of Hamming loss, Schapire and Singer [9] have extended Adaboost to a multi-class multi-label version. Denote the sample space by $\mathcal{X}$ and the label set by $\mathcal{Y}$. A sample of a multi-class multi-label problem is a pair $(\mathbf{x}, Y)$, where $\mathbf{x} \in \mathcal{X}, Y \subseteq \mathcal{Y}$. For $Y \subseteq \mathcal{Y}$, define Y[l] for $l \in \mathcal{Y}$ as

$$Y[l] = \begin{cases} 1 & \text{if } l \in Y \\ -1 & \text{if } l \notin Y \end{cases} \qquad (1)$$

Fig. 2 lists the details of the multi-class Adaboost algorithm.

Given: the sample set $S = \{(\mathbf{x}_1, Y_1), \ldots, (\mathbf{x}_m, Y_m)\}$, and the size of the final strong classifier T.

Initialize: $D_1(i,l) = 1/(mk)$, i=1..m, l=1..k, where $k = |\mathcal{Y}|$.

For t = 1… T

- Under the distribution D_t, select a weak classifier $h_t : \mathcal{X} \times \mathcal{Y} \to [-1,1]$ from the weak classifier pool to maximize the absolute value of

$$r_t = \sum_{i,l} D_t(i,l) Y_i(l) h_t(\mathbf{x}_i, l). \qquad (2)$$

- Let $$\alpha_t = \frac{1}{2} \ln\left(\frac{1+r_t}{1-r_t}\right). \qquad (3)$$

- Update the distribution:

$$D_{t+1}(i,l) = \frac{D_t(i,l)\exp\left(-\alpha_t Y_i[l] h_t(\mathbf{x}_i, l)\right)}{Z_t}, \qquad (4)$$

Where Z_t is a normalization factor so that D_{t+1} is a p.d.f.

Output the final strong classifier:

$$H(\mathbf{x}, l) = \mathrm{sign}\left(\sum_t \alpha_t h_t(\mathbf{x}, l)\right) \qquad (5)$$

Figure 2: The continuous multi-class Adaboost algorithm.

From Eq.5, it can be seen a sample can belong to several classes. To get a multi-class single-label classifier, the Eq.5 is adjusted to the following form

$$H(\mathbf{x}) = \arg\max_l \left(\sum_t \alpha_t h_t(\mathbf{x}, l)\right) \qquad (6)$$

and furthermore a confidence of H is defined as

$$Conf_H(\mathbf{x}) = \left| \frac{\sum_t \alpha_t h_t(\mathbf{x}, H(\mathbf{x}))}{\sum_t \alpha_t} \right|. \qquad (7)$$

4. Haar Feature Based LUT Weak Classifiers

In order to use the continuous multi-class Adaboost algorithm, a weak classifier pool of simple features should be configured. We construct a weak classifier based on the Haar feature, which is a kind of simple rectangle feature proposed by Viola and Jones [12], and can be calculated very fast through the integral image. For each Haar feature, one weak classifier is configured. In [12], a threshold-type weak classifier, whose output is Boolean value, is used, that is, $h(\mathbf{x}) = \mathrm{sign}\left[f_{Haar}(\mathbf{x}) - b\right]$, where f_{Haar} is the Haar feature and b is a threshold, see Fig.3a. The main disadvantage of this threshold model is that it is too simple to fit complex distributions, such as a multi-Gaussian model. For a multi-class problem, $\varpi_1, \ldots, \varpi_k$, we propose a real-valued 2D Look-Up-Table (LUT) type weak classifier, see Fig.3b. Assuming f_{Haar} has been normalized to [0, 1], the range is divided into n sub-ranges:

$$bin_j = [(j\text{-}1)/n,\ j/n),\ j = 1, \ldots, n \qquad (7)$$

The weak classifier $h(\mathbf{x}, l)$ is defined as

$$\text{If } f_{Haar}(\mathbf{x}) \in bin_j \text{ then } h(\mathbf{x}, l) = 2P_l^{(j)} - 1, \qquad (8)$$

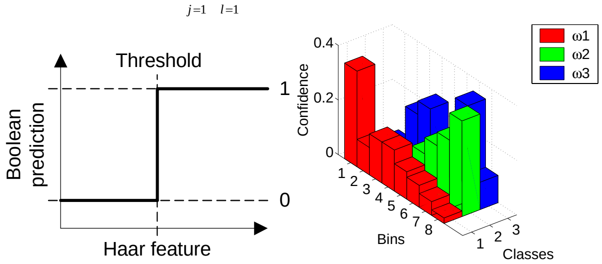
where $P_l^{(j)} = P(\mathbf{x} \in \varpi_l \mid f_{Haar}(\mathbf{x}) \in bin_j)$.

Given the characteristic function

$$B_n^{j,l}(u, y) = \begin{cases} 1 & u \in \left[(j-1)/n,\ j/n\right) \wedge y = l \\ 0 & \text{otherwise} \end{cases}, \qquad (9)$$

The LUT weak classifier can be formally expressed as:

$$h_{LUT}(\mathbf{x}, y) = \sum_{j=1}^{n} \sum_{l=1}^{k} \left(2P_l^{(j)} - 1\right) B_n^{j,l}\left(f_{Haar}(\mathbf{x}), y\right). \qquad (10)$$



(a) Threshold (b) LUT

Figure 3: Threshold week classifier vs. LUT week classifier.

5. Experiments and Results

Our facial expression recognition system is based on the above learning method with Haar feature based LUT weak classifiers. In this paper, the expression recognition is defined as a seven-class problem of anger, disgust, fear, happiness, neutral, sadness and surprise. As mentioned in Section 1, we use the JAFFE database [1] [14] to train an expression classifier. The database is a set of 213 images of 7 different expressions posed by 10 Japanese females. In preprocessing, all the face samples in JAFFE database are normalized to 36×36-pixel patches and processed by histogram equalization, see Fig.4 for some samples. For generality, no color information is used. Furthermore, in order to enhance the generalization performance of Adaboost learning, we define following transformations: (1) mirror reflection, (2) inflation by a ratio of 1.1, (3) stretch by a ratio of 1.1, (4) rotation an angle $\pm 5°$, and use those transformations and their combinations to create more pseudo samples. So in total, we have 5112 face samples for training an expression classifier.

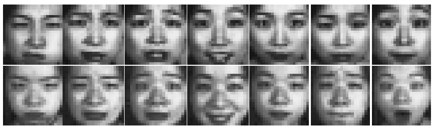


Figure 4: Normalized expression sample. From left to right: anger, disgust, fear, happiness, neutral, sadness and surprise.

For training the expression classifier, we set k=7 and T = 1,400 in the continuous multi-class Adaboost algorithm (Fig.2), i.e., seven categories of expression to be discriminated and 1,400 weak classifiers to be selected to construct one strong classifier. In each training round, the parameter of each classifier is optimized and the feature that provided best performance on the boosted distribution is chosen. For comparison, a classifier based on SVMs with RBF-kernel is also trained on the same training set.

Table 1: Experimental results on a database of 206 images with 385 frontal facial expressions

	average correct rate of 7 categories	average processing time per image
Our Method	92.4%	0.11 ms
SVMs	91.6%	28.7 ms

With our expression classifier learned, we test it first on the JAFFE database. Because these data belong to the training set, it achieves a high correct rate up to 98.9%. Then, we test it on an image database independent of training set which consists of images from World Wide Web and images captured by cameras in our laboratory. This database has 206 images that contain 385 frontal facial expressions. Experimental results on this test set are listed in Table 1 and in Fig. 5. It can be seen that the correctness of our method is slightly better than that of the SVMs method. However in average, our method takes only 0.11ms to process a face sample on a Pentium IV 2.53G Hz PC which is nearly 300 times faster than the SVMs method.

Figure 5: Experimental results on a database of 206 images with 385 frontal facial expressions.

Some of the failures found are perhaps due to too lower resolutions of face images (Fig.6d), and some facial expressions are in fact too difficult to discriminate due to the fineness variation and complexity of facial expression (Fig.6a, b). In fact those expressions misclassified by the expression classifier (Fig.6b, c, d) are difficult for us human beings to discriminate too. The labeled expressions in Fig. 6 are happiness, surprise, neutral and surprise from left to right; but they are recognized as happiness, happiness, disgust and fear respectively.

	a	b	c	d
labeled：	happiness	surprise	neutral	surprise
recognized:	happiness	happiness	disgust	fear

Figure 6: Some complex facial expressions.

For potential real applications, we have integrated the expression recognition module with an automatic face detection and facial feature landmarks extraction system as illustrated in Fig.1. The input of this system is real-life photos and output is some possible face blocks labeled with the category of expression. Fig.7 shows some results on real-life photos.

6. Conclusion

In this paper we have proposed a novel facial expression recognition method in which the expression classifier is learned by boosting Haar feature based Look-

Up-Table weak classifiers. Experimental results on a database of 206 images with 385 frontal facial expressions show that the boosting method has achieved a better performance than SVMs method, especially in speed. Furthermore we have developed an automatic real time facial expression recognition system that may be useful in future humanoid toy robot. Although this paper focuses on facial expression recognition problems, our framework can be used in other two-class or multi-class pattern recognition problems too.

7. Acknowledgement

This work is supported by National Science Foundation of China under grant No.60332010. We would like to thank M.J. Lyons for kindly providing facial expression database used in this paper.

8. References

[1] The Japanese Female Facial Expression (JAFFE) Database, http://www.mis.atr.co.jp/~mlyons/jaffe.html.

[2] Maja Pantic, and Leon J.M. Rothkrantz, "Automatic Analysis of Facial Expressions: The State of Art", *IEEE Trans. on PAMI*, VOL. 22, NO. 12, 2000, pp. 1424-1445.

[3] M.J. Lyons, J. Budynek, and S. Akamatsu, "Automatic Classification of Single Facial Images", *IEEE Trans. on PAMI*, VOL. 21, NO. 12, 1999, pp.1357-1362.

[4] C. Padgett and G.W. Cottrell, "Representing Face Images for Emotion Classification", *in Proc. Conf. Advances in Neural Information Processing Systems*, Denver, USA, 1996, pp. 894-900.

[5] P. Ekman and W.V. Friesen, *Unmasking the Face*, Prentice Hall, New Jersey, 1975.

[6] T. Otsuka, and J. Ohya, "Spotting Segments Displaying Facial Expressions from Image Sequences Using HMM", *in IEEE Proc. Int'l Conf. on FG*, Nara, Japan, 1998, pp. 442-447.

[7] M. S. Bartlett, G. Littlewort, I. Fasel, et al., "Real Time Face Detection and Facial Expression Recognition: Development and Applications to Human Computer Interaction", *in IEEE CVPR Workshop on Computer Vision and Pattern Recognition for Human-Computer Interaction*, Wisconsin, USA, 2003.

[8] T. Kanade, J. F. Cohn, and Y. L. Tian, "A comprehensive database for facial expression analysis", *in IEEE Proc. Int'l Conf. on FG*, Grenoble, France, 2000, pp 46-53.

[9] R. E. Schapire and Y. Singer, "Improved boosting algorithms using confidence-rated predictions", *Machine Learning*, VOL. 37, NO. 3, 1999, pp. 297-336.

[10] G. Shakhnarovich, P. Viola and B. Moghaddam, "A Unified Learning Framework for Real Time Face Detection and Classification", *in IEEE Proc. Int'l Conf. on FG*, Grenoble, France, 2000, pp. 14-21.

[11] Bo Wu, Haizhou Ai, Chang Huang, "LUT-Based Adaboost for Gender Classification", *in Int'l Conf. on Audio- and Video-Based Biometric Person Authentication*, Guildford, UK, 2003, pp. 104-110.

[12] P. Viola and M. Jones, "Rapid Object Detection using a Boosted Cascade of Simple Features", *in IEEE Proc. Int'l Conf. on CVPR*, Hawaii, USA, 2001, pp. 511-518.

[13] Tong WANG, Haizhou AI, Gaofeng HUANG, A Two-Stage Approach to Automatic Face Alignment, *in Proc. SPIE International Symposium on Multispectral Image Processing and Pattern Recognition*, Beijing, China, 2003, pp. 558-563.

[14] M. J. Lyons, S. Akamatsu, M. Kamachi, et al., "Coding Facial Expressions with Gabor Wavelets", *in IEEE Proc. Int'l Conf. on FG*, Nara, Japan, 1998, pp. 200-205.

Figure 7: Expression recognition results on real-life photos.