Development of an Intelligent Facial Expression Recognizer for Mobile Applications

Siu-Yeung Cho, Teik-Toe Teoh, and Yok-Yen Nguwi

Abstract. In the light of fast pace smart phone development, increasing numbers of applications are being developed to cater for portability. A real-time facial expression recognition application is develop that was tested in Windows Mobile environment. The underlying algorithm adopted in this work uses Boosting Naïve Bayesian (BNB) approach for recognition. We examine the structure of training data and the effect of attributes on the class probabilities through the use of Naïve Bayesian classifier (NBC). The experiments carried out show that we have achieved the important features of mobile application: speed and effi-ciency. This work is believed to be the first recorded initiative that de-ploys facial expression recognition into a mobile phone. It seeks to pro-vide a launching point for a sound and portable mobile application that is capable of recognizing different facial expressions.

1 Introduction

To date, facial expression recognition system becomes popular in many application domains. It can be applied in, for examples, when a facial expression recognition system is installed in a robot, it can communicate with human better. Robot/Human interaction is gaining focus nowadays. This contributes to the field of human computer interaction which will have great impact on the society following the speed advancement of robot. Innovation like household robot which can en-courage the owner when he is feeling down, comfort a child when he is crying, making a joke when the owner is bored and etc. The same concept can be further developed into an intelligent house that understands people's emotion through his face expression and entertain him. On the other hand, another possible application can be explored in an intelligent portable device for social communication. Eight out of ten people in Singapore own a mobile phone, be it a simple mobile phone or a high end 3G handheld device. Mobile phone becomes part of our lives that most people cannot live without. The development of mobile phone

---

Siu-Yeung Cho, Teik-Toe Teoh, and Yok-Yen Nguwi
School of Computer Engineering,
Nanyang Technological University
Singapore 639798
e-mail: davidcho@pmail.ntu.edu.sg

K. Nakamatsu et al. (Eds.): New Advan. in Intel. Decision Techno., SCI 199, pp. 21–29.
springerlink.com

hardware and soft-ware development becomes the hot pies in the commercial market. The commonly used smart phones are Windows Mobile Personal Digital Assistant (PDA), Black-Berry, Apply iPhone, Palm OS handhelds, Symbian OS handhelds and etc. Among the smart handheld devices, Windows Mobile platform is the most commonly used one.

The motivation behind this work is that we aim to develop a sound portable device that can help autistic children to understand the emotions of surrounding people. Emotion is a state of feeling involving thoughts, physiological changes, and an outward expression. There are five theories which attempt to understand the sequence of processes that we are experiencing when we are feeling a certain type of emotion. They are James-Lange theory, Cannon-Bard theory, Lazarus theory, Schachter-Singer theory, and Facial Feedback theory [1].

According to the facial feedback theory [1], emotion is the experience of changes in our facial muscles. In other words, when we smile, we experience pleasure or happiness. When we frown, we experience sadness. It is the changes in our facial muscles that direct our brains and provide the basis for our emotions. As there are many possibilities of muscle configurations in our face, there is seemingly unlimited number of emotions. Facial expression recognition is a challenging task. The challenges of such system are light variation, direction of subject face, the quality of image acquisition device, and occlusion problem. Recent research works have been done to contribute in this area, for examples, in [2-5].

This work attempts to use Naïve Bayes Classifier (NBC) to classify facial emotions. Bayesian classifier is the most popular classifier among the probabilistic classifiers used in the machine learning community and is often used for benchmarking. Naive Bayesian classifier is perhaps one of the simplest yet powerful techniques in constructing predictive models from labeled training sets. NBC can also provide valuable insights of the training data by exposing the relations between attribute values and classes. The NBC is robust and often outperform other more complex machine learning methods.

In this paper, we propose an approach to recognize emotion from facial images through the use of Boosting Naïve Bayesian (BNB) approach. The observation obtained through the classification process will be discussed. Four classes of facial emotion are investigated; they are namely neutral, joy, sad and surprise as shown in Fig 1. The focus of this work is to show how the BNB approach can be used to segregate the four classes of the facial emotion in a mobile application.

The paper is organized as follows: Section 2 provides an overview of our proposed system. Section 3 presents the experimental results and analysis. Finally, conclusion of this paper is drawn in Section 4.

Fig. 1 Four categories of facial expressions. (a) Neutral, (b) Joy, (c) Sad, and (d) Surprise

2 Our System

In this section, we outline the structure of facial expression recognizer. The proposed model is going to recognize four types of facial expression, namely, neutral, joy, sad and surprise. A probabilistic based approach using the Naïve Bayesian Boost is adopted to recognize these four types of facial expressions.

Figure 2 displays the block diagram of the entire system. The system is com-posed of three major blocks, the feature locator, the feature extractor, and the classifier. The feature locator finds crucial fiducial points for subsequent feature ex-traction processing. We adopted Gabor features and Gini extractions which will be discussed in the following sub-sections. Finally the meaningful features are classified into the corresponding class.

The fiducial points are located at the eyes, nose, and mouth. The fiducial points define the extended feature components. The component-based feature detector has two levels, namely, the micro SVM based independent component detector and the macro SVM based independent component detector. The micro level uses linear SVM based independent component detection. Each component classifier was trained on a set of extracted facial components (the 4 key fiducial components) and on a set of randomly selected non-face patterns. The macro level uses the maximum outputs of the component classifiers within rectangular search regions as inputs to a combination SVM classifier. The macro SVM performs the final detection of the face component regions.

The feature extractor adopted Gabor wavelet feature extraction incorporating with Gini-based feature selection. An image is first convoluted with Gabor wavelet filters so as to extract the facial features. We have chosen 7 orientations and 2 spatial frequencies; generating a total of 14 Gabor filtered images. The frequency selected is from 0.4 to 0.8, whereas the orientations are in the multiples of π/7 ranging from 0 to π. Gabor wavelet constructed 61 regions of 14 different features for each input data. The feature extraction process starts by marking the left eye, right eye and mouth positions. With the 854 (61x14) features obtained, we make

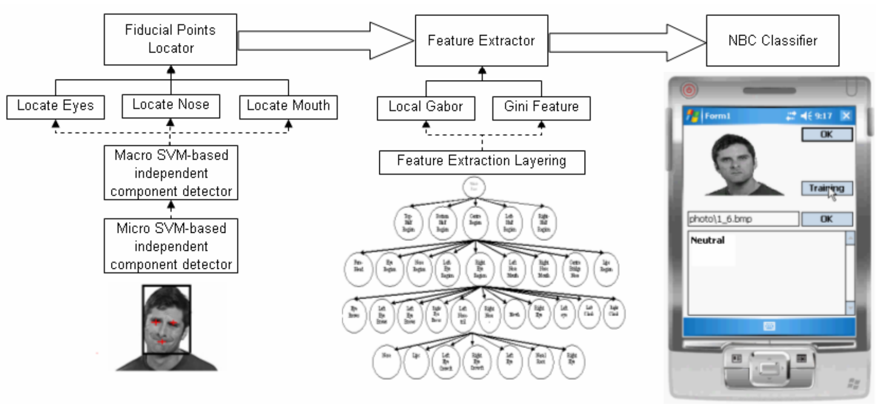


Fig. 2 The system block diagram

use of Gini to shrink down to ten features in which they are the best features to describe the face. The final stage is the probabilistic-based classification performed by the NBC classifier.

2.1 Gabor Feature Extraction

Gabor wavelet is a popular choice because of its capability to mimic mammals' visual cortex. The primary cortex of human brain interprets visual signals. It consists of neurons, which respond differently to different stimuli attributes. The receptive field of cortical cell consists of a central ON region surrounded by 2 OFF regions, each region elongated along a preferred orientation [6]. According to Jones and Palmer, these receptive fields can be reproduced fairly well using Daugman's Gabor function [7]. There is considerable evidence that the parameterized family of 2-D Gabor filters, proposed by Daugman in 1980, suitably models the profile of receptive cells in the primary visual cortex. Gabor filters models the properties of spatial localization, orientation selectivity, and spatial frequency selectivity and phase relationship of the receptive cells [8].

The Gabor wavelet function can be represented by:

$$g(x, y) = g_1(x, y)\exp(j2\pi Wx)\,, \tag{2.1}$$

where

$$g_1(x, y) = \left(\frac{1}{2\pi\sigma_x\sigma_y}\right)\exp\left(-\frac{1}{2}\left(\frac{x^2}{\sigma_x^{\ 2}} + \frac{y^2}{\sigma_y^{\ 2}}\right)\right). \tag{2.2}$$

We consider that the receptive field (RF) of each cortical cell consists of a central ON region (a region excited by light) surrounded by two lateral OFF regions (excited by darkness) [18]. Spatial frequency W determines the width of the ON and OFF regions. σ_x^2 and σ_y^2 are spatial variances which establish the dimension of the RF in the preferred and non-preferred orientations.

2.2 Gini Index Feature Reduction and Selection

Gini Index selects features based on information theories [9, 10]. It measures the impurity for a group of labels. Gini Index for a given set s of points assigned to two classes C_1 and C_2 is given below:

$$GINI(s) = 1 - \sum_{j=1,2}\left[p\left(C_j \mid s\right)\right]^2. \tag{2.3}$$

With $p\left(C_j \mid s\right)$ corresponds to the frequency of class C_j at set *S*. The maximum $1-\frac{1}{nc}$ occurs when points are equally distributed among all classes, which implies less interesting information. On the other hand, the minimum 0.0 occurs when all points belong to one class which represents the most interesting information. We then sort the *n* features over different classes of samples in ascending order based on their best Gini index. Low Gini index corresponds to high ranking discriminative features.

2.3 Probabilistic-Based Classification

A Naive Bayesian (NB) classifier is a simple probabilistic classifier base on Bayesian' theorem with strong (Naive) independence assumptions [11, 12]. The NB classifiers often work much better in many complex real-world problems. The classifier requires only a small amount of training data to estimate the parameters necessary for classification. The independence assumption in Naïve Bayesian may lead to some unexpected results in the calculation of posteriori probability. The NB classifier has several attractive properties like the decoupling of the class conditional feature distributions which helps to alleviate problems stemming from the curse of dimensionality. The data arrives at the correct classification as long as the correct class is more probable than any other class. In overall, the classifier is robust enough to ignore serious deficiencies in the underlying Naive probability model.

Boosting Naïve Bayesian Algorithm:

GINI feature selections:

1. Assign each training sample with weight=1.
2. For ten iteration (ten features):
   - Sort features index S.
   - Split S.
   - Break if GINI criterion is satisfied.

BNB classification:

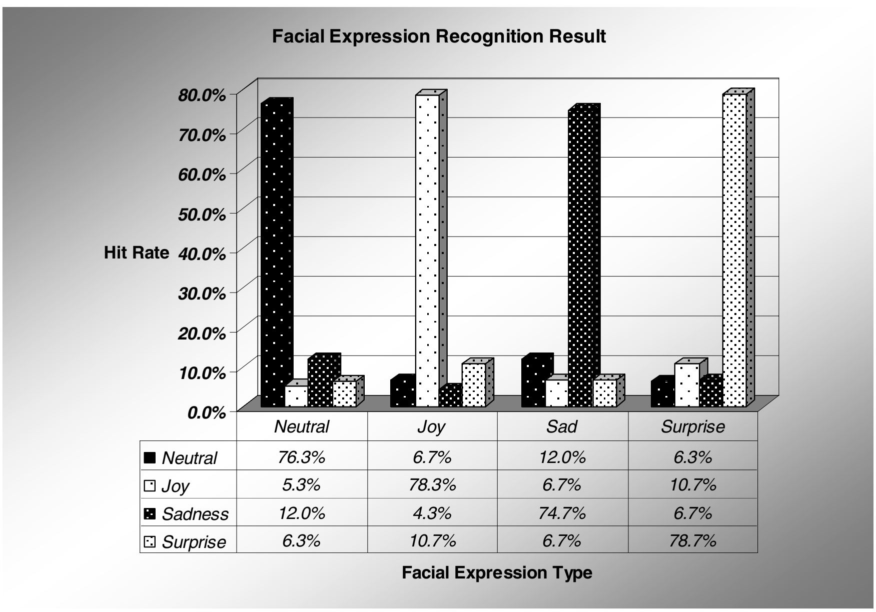
1. Apply simple Bayesian to weighted data set.
2. Compute error rate.
3. Iterate the training examples.
   - Multiply the weight by $\frac{e}{1-e}$.
   - Normalize the weight
4. Add $-\log\frac{e}{1-e}$ to weight of class predicted
5. Return class with highest sum

Fig. 3 The proposed Naïve Bayesian with boosting algorithm

The proposed algorithm adopted the Naïve Bayesian with boosting. Each iteration of boosting uses the Gini reduction and selection method and to remove redundant features. The main reason for using boosting NB approach is that the embedded feature selection technique makes the data more suitable for classification. The algorithm is summarized as shown in Fig. 3.

3 Experimental Results and Analysis

In this section, we will assess the ability of the system to recognize different facial expressions. We have adopted the Mind Reading DVD [13], a computer-based guide to emotions, developed by a team of psychologists led by Prof. Simon Baron-Cohen at the Autism Research Centre, University of Cambridge. The database contains images of approximately 100 subjects. Facial images are of size 320x240 pixels, 8-bit precision grayscale in PNG format. Subjects' age ranges from 18 to 30. Sixty-five percent were female, 15 percent were African-American, and three percent were Asian or Latino. Subjects were instructed by experimenter to perform a series of facial displays. Subjects began each display from a neutral face. Before performing each display, the experimenter described and modeled the desired display. The model recognizes four types of facial expression: neutral, joy, sadness and surprise. Twenty images were used for training, 5 images for each emotion.



	Neutral	Joy	Sad	Surprise
Neutral	76.3%	6.7%	12.0%	6.3%
Joy	5.3%	78.3%	6.7%	10.7%
Sadness	12.0%	4.3%	74.7%	6.7%
Surprise	6.3%	10.7%	6.7%	78.7%

Fig. 4 Facial Expression Recognition Result of the System

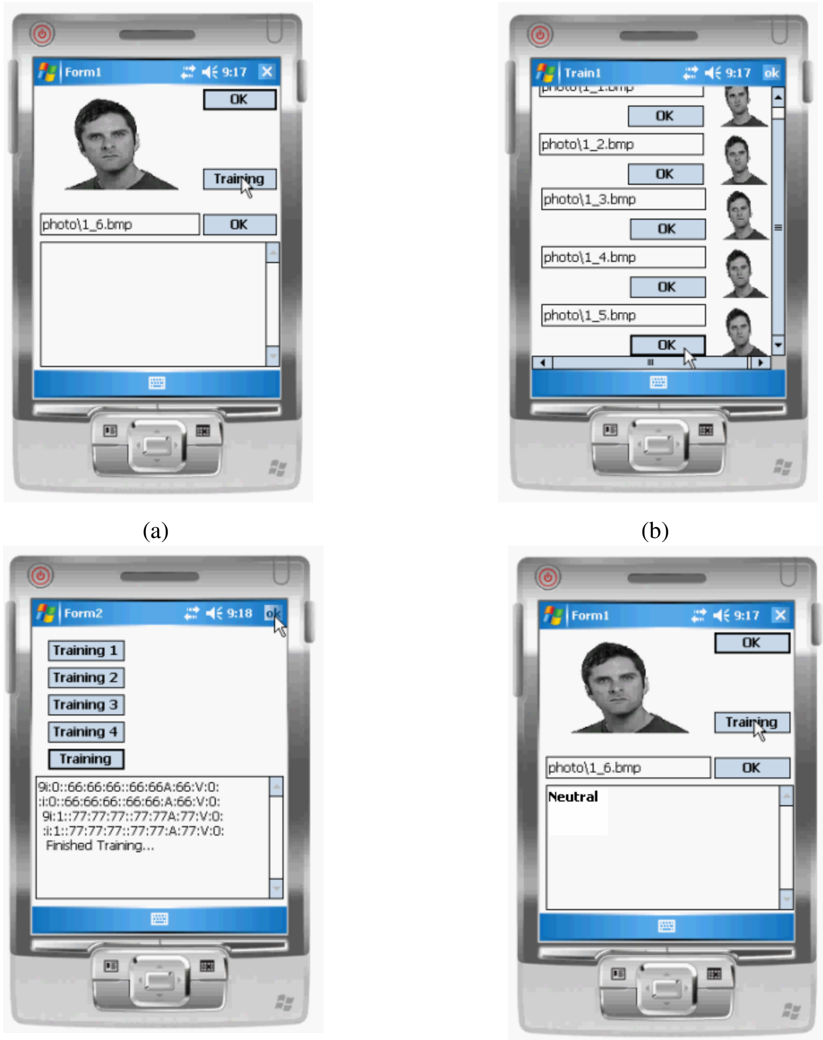
The facial expression recognition result is shown in Fig. 4. The confusion matrix is included in the figure as well where the column of the matrix represents the instances in a predicted class, in which each row represents the instances in an actual class. The system correctly recognizes 76.3% of neutral, 78.3% of joy, 74.7% of sad and 78.7% of surprise expressions amongst 100 subjects in the database, although some facial expressions do get confused by the wrong class, however at an acceptable range of less than 12%.

In addition, comparisons with other approaches are necessary for us to investigate how the recognition performance of our approach can be benchmarked with others. Table 1 shows the recognition results for facial expression recognition using T-test, Euclidean, and K-nearest neighbour approaches. Only the average and maximum hit rates are included in the table. According to the results in the table, our approach achieves the most optimal result. The T-test assesses whether the means of different groups are statistically different from each other. K-nearest neighbour algorithm is a method for classifying objects based on closest training examples in the feature space. These approaches are generally used for benchmarking. Our approach that combines Gini and Boosting Naïve Bayesian achieves average of 75% and highest of 100% outperforms the rest.

The application is deployed in a Hewlett-Packard's IPAQ PDA phone with Windows Mobile operating system. Fig. 5 shows the screen capture of the system. Fig. 5(a) shows the main menu that provides selection for training and testing. The training can be done on the smart phone by loading different training images as shown in Fig. 4(b). The 4 classes of facial expressions are trained using the GUI as shown in Fig. 4(c). Lastly we can load the corresponding image for testing Fig. 4(d). The average recognition time is about 5 seconds.

Table 1 Result benchmarking with T-test, Euclidean and KNN approaches

Feature selection	Classifier	Average	Maximum
Gini	Naïve Bayesian	75%	100%
Gini	Euclidean	63%	83%
Gini	KNN	57%	99%
T-test	Euclidean	59%	69%
T-test	KNN	59%	89%
T-test	Naïve Bayesian	57%	100%
Euclidean	Euclidean	58%	69%
Euclidean	KNN	57%	100%
Euclidean	Naïve Bayesian	53%	99%



(a) (b) (c) (d)

Fig. 5 Screenshot of facial expression recognizer in PDA. (a) Main menu. (b) Image loader (c) Training GUI (d) Test result

4 Conclusion

A real-time facial expression recognition system is presented in this paper. The system consists of fiduciary points localizer, features extractor, and the classifier. The fiduciary points localizer consists of both micro and macro SVM based independent component detectors. The feature extractor processes fiduciary points with Gabor wavelet convolution and Gini index sorting. The classifier segregates the facial expression based on Boosting Naïve Bayesian (BNB) algorithm.

The system is able to recognize four classes of facial expression: neutral, joy sad, and surprise. The experimental result shows that Gini indexing to the features plays a very important role in recognizing different facial emotion. The experimental result also suggests the combination of Gini and BNB achieves the most optimal result. The average hit rate is around 75%. The application is deployed in a Hewlett-Packard's IPAQ PDA phone. Various facial expressions can be trained in the smart phones and gets recognition within 5 seconds.

Currently, we are working on the improvement of the BNB classification in its structure and algorithm to obtain better accuracy and reduce the computation time for training. We are also looking into further enhance the system to obtain live images through the built-in camera in smart phones.

References

[1] Heffner, C.L.: Psychology 101, The virtual psychology classroom (2001), `http://allpsych.com/psychology101/emotion.html`

[2] Ji, Y.Z.Q.: Active and dynamic information fusion for facial expression understanding from image sequences. IEEE Transactions on Pattern Analysis and Machine Intelligence 27(5), 699–714 (2005)

[3] Wu, Y., Liu, H., Zha, H.: Modeling facial expression space for recognition. In: The IEEE/RSJ International Conference on Intelligent Robots and Systems (2005)

[4] Zeng, Z., Fu, Y., Roisman, G.I., Wen, Z., Hu, Y., Huang, T.S.: One-class classification for spontaneous facial expression analysis. In: The International Conference on Automatic Face and Gesture Recognition (2006)

[5] Xiang, T., Leung, M.K.H., Cho, S.Y.: Expression recognition using fuzzy spatio-temporal modeling. Pattern Recognition 41(1), 204–216 (2007)

[6] Jones, J.P., Palmer, L.A.: An evaluation of the Two-Dimensional Gabor Filter model of simple Receptive fields in cat striate cortex. J. Neurophysiol. 58(6), 1233–1258 (1987)

[7] Daugman, J.: Uncertainty relation for resolution in space, spatial frequency, and orientation optimized by two-dimensional visual cortical filters. J. Opt. Soc. Amer. 2(7), 1160–1169 (1985)

[8] La Cara, G.E., Ursino, M., Bettini, M.: Extraction of Salient Contours in Primary Visual Cortex: A Neural Network Model Based on Physiological Knowledge. In: Engineering in Medicine and Biology Society, 2003. Proceedings of the 25th Annual International Conference of the IEEE, September 17-21, 2003, vol. 3, pp. 2242–2245 (2003)

[9] Hall, M.A., Smith, L.A.: Feature Selection for Machine Learning: Comparing a Correlation-Based Filter Approach to the Wrapper. In: FLAIRS Conference, pp. 235–239 (1999)

[10] Zhou, X.J., Dillion, T.S.: A Heuristic - Statistical Feature Selection Criterion For Inductive Machine Learning In The Real World. In: Proceedings of the 1988 IEEE International Conference on Systems, Man, and Cybernetics, August 1988, vol. 1, pp. 548–552 (1988)

[11] Friedman, N., Geiger, D., Goldszmidt, M.: Bayesian network classifiers. Machine Learning 29, 131–163 (1997)

[12] Rish, I.: An empirical study of the naïve Bayes classifier. Technical Report RC 22230 (2001)

[13] Baron-Cohen, S., Golan, O., Wheelwright, S., Hill, J.J.: Mind Reading: The Interactive Guide to Emotions. Jessica Kingsley Publishers, London (2004)