Facial Expression Recognition with Local Binary Patterns and Linear Programming[1]

X. Feng*, **, M. Pietikäinen*, and A. Hadid*

**Machine Vision Group, Infotech Oulu and Department of Electrical and Information Engineering, P. O. Box 4500 FI-90014 University of Oulu, Finland*

e-mail: {xiaoyi, mkp, hadid}@ee.oulu.fi

***College of Electronics and Information, Northwestern Polytechnic University, 710072 Xi'an, China*

e-mail: fengxiao@nwpu.edu.cn

Abstract—In this work, we propose a novel approach to recognize facial expressions from static images. First, the local binary patterns (LBP) are used to efficiently represent the facial images and then the linear programming (LP) technique is adopted to classify seven facial expressions—anger, disgust, fear, happiness, sadness, surprise, and neutral. Experimental results demonstrate an average recognition accuracy of 93.8% in the JAFFE database, which outperforms the rates of all other reported methods in the same database.

INTRODUCTION

Facial expression recognition from static images is a more challenging problem than from image sequences because less information for expression actions is available. However, information in a single image is sometimes enough for expression recognition, and in many applications it is also useful to recognize single image's facial expression.

In the recent years, numerous approaches to facial expression analysis from static images have been proposed [1, 2]. These methods differ generally in face representation and similarity measure. For instance, Zhang [3] used two types of features: the geometric position of 34 manually selected fiducial points and a set of Gabor wavelet coefficients at these points. These two types of features were used both independently and jointly with a multilayer perceptron for classification. Guo and Dyer [4] also adopted a similar face representation, combined with the linear programming (LP) technique to carry out simultaneous feature selection and classifier training, and they reported a better result. Lyons *et al.* used a similar face representation with a simple LDA-based classification scheme [5]. All the above methods required the manual selection of fiducial points. Buciu *et al.* used ICA and Gabor representation for facial expression recognition and reported a good result in the same database [6]. However, a suitable combination of feature extraction and classification is still an imperative question for expression recognition.

In this paper, we propose a novel method for facial expression recognition. In the feature extraction step, the local binary pattern (LBP) operator is used to describe facial expressions. In the classification step, seven expressions (anger, disgust, fear, happiness, sadness, surprise, and neutral) are decomposed into 21 expression pairs such as anger-fear, happiness-sadness, etc. Twenty-one classifiers are produced by the LP technique, each corresponding to one of the 21 expression pairs. A simple binary tree tournament scheme with pairwise comparisons is used for classifying unknown expressions.

FACE REPRESENTATION WITH LOCAL BINARY PATTERNS

Figure 1 shows the basic LBP operator [7], in which the original 3×3 neighborhood at the left is thresholded by the value of the center pixel and a binary pattern code is produced. The LBP code of the center pixel in the neighborhood is obtained by converting the binary code into a decimal one.

Based on the operator, each pixel of an image is labeled with an LBP code. The 256-bin histogram of the labels contains the density of each label and can be used as a texture descriptor of the considered region.

Feature extraction is implemented as follows: First, the face image is divided into several nonoverlapping blocks (10×8 in our method after experimenting with different block sizes). Then, LBP histograms are calcu-

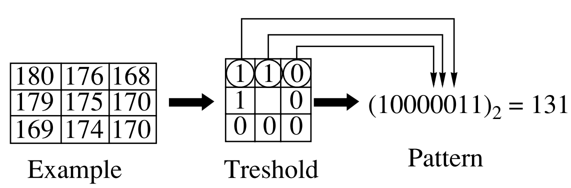

Fig. 1. The basic LBP operator.

[1] The text was submitted by the authors in English.

Received October 25, 2004

Pattern Recognition and Image Analysis, Vol. 15, No. 2, 2005, pp. 546–548.

lated for each block. Finally, the block LBP histograms are concatenated into a single vector. As a result, the facial expressions are represented by the LBP and the shape of the face is recovered by the concatenation of different local histograms.

To reduce the length of the feature vector, all the training samples' feature vectors are summed up and then averaged by the number of training samples. Then, for each region of an image, the patterns whose occurrence frequency is lower than a threshold are discarded.

EXPRESSION CLASSIFICATION BASED ON LINEAR PROGRAMMING TECHNIQUE

In [8], a single linear programming (LP) formulation is proposed which generates a plane that minimizes an average sum of misclassified points belonging to two disjoint points set. Consider the two point-sets A and B in the n-dimensional space R^n represented by the $m \times n$ matrix A and the $k \times n$ matrix B. The separating plane is as follows:

$$P := \{x | x \in R^n, x^T \omega = \gamma\}. \tag{1}$$

Here $\omega \in R^n$ is normal to the separating plane with a distance $|\gamma|/\|\omega\|$ to the origin.

The separating plane P determines two open half-spaces, $\{x|x \in R^n, x^T\omega > \gamma\}$ containing mostly points belonging to A, and $\{x|x \in R^n, x^T\omega < \gamma\}$ containing mostly points belonging to B. That is, we wish to satisfy

$$A\omega > e\gamma, \quad B\omega < e\gamma. \tag{2}$$

Here e is a vector of all 1's with the appropriate dimension. To the extent possible, or upon normalization,

$$A\omega \geq e\gamma + e, \quad B\omega \leq e\gamma - e. \tag{3}$$

Conditions (2) or (3) can be satisfied if and only if A and B do not intersect, which in general is not the case. We thus attempt to satisfy (3) by minimizing some norm of the average violations of (3),

$$\min_{\omega, \gamma} \frac{1}{m}\|(-A\omega + e\gamma + e)_+\|_1 + \frac{1}{k}\|(B\omega - e\gamma + e)_+\|_1. \tag{4}$$

Here x_+ denotes the vector in R^n satisfying $(x_+)_i := \max\{x_i, 0\}$, $i = 1, 2, \ldots n$. The norm $\|\cdot\|_p$ denotes the p norm, $1 \leq p \leq \infty$.

Formulation (4) is equivalent to the following robust linear programming formulation

$$\min_{\omega, \gamma, y, z} \frac{e^T y}{m} + \frac{e^T z}{k} \tag{5}$$

subject to

$$\begin{cases} -A\omega + e\gamma + e \leq y \\ B\omega - e\gamma + e \leq z \\ y \geq 0, z \geq 0. \end{cases}$$

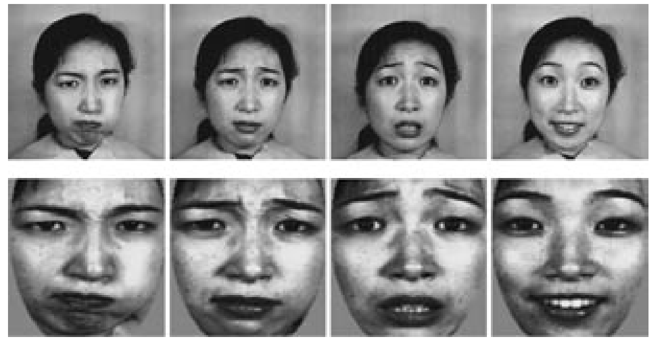

Fig. 2. Samples from original Japanese Female Facial Expression images before and after preprocessing.

In our study, we use formulation (5) as a classifier to minimize wrong classifications.

Since formulation (5) is only used for separating two sets points, the seven-expression classification problem is decomposed into 21 two-class problems (anger-disgust, anger-fear, disgust-fear, etc.). In the training stage, 21 classifiers corresponding to 21 expression pairs are formed with 21 pairs of $\{\omega, \gamma\}$. In the testing stage, feature vector of a testing sample is fed to the different classifiers for recognition.

EXPERIMENTAL RESULTS

To assess the validity and efficiency of our approach, we experimented with the Japanese Female Facial Expression (JAFFE) database [9]. The database contains 213 images in which ten persons are expressing three or four times the seven basic expressions.

First, we preprocess the images using the subsystem of the CSU Face Identification Evaluation System [10]. The images are registered using eye coordinates and cropped with an elliptical mask to exclude nonface area from the image. As a result, the size of each preprocessed image is 150×128. Sample images from the original and the preprocessed images are shown in Fig. 2.

We divide the database into ten roughly equally sized sets: nine sets are used for training and the remaining one for testing. The above process is repeated so that each of the ten roughly equally sized sets is used once as the test set. The average result over all ten cycles is considered as the recognition rate of one trial.

We repeat the above procedure 20 times and get 20 average results for these trials, of which the highest rate is 96.3%, the lowest one is 92.0%, the mean is 93.8%, and the standard deviation is 1.35%.

In [3], a multilayer perceptron was used with 90.1% recognition accuracy. In [4], a technique called feature selection via linear programming was used and they achieved an accuracy of 91%. Therefore, better results are obtained with our approach.

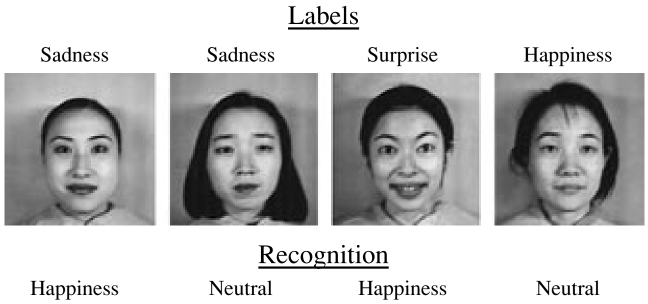

Fig. 3. Examples of disagreement.

It was reported in [3] that the expressers found it most difficult to pose fear expressions accurately and a human has more difficulty in recognizing fear. They achieved a recognition rate of 92.3% when all fear images were excluded. In our experiments, a highest rate of 95.6%, a lowest rate of 93.4% and an average rate of 94.6% are obtained when all fear images are excluded.

In [5], a recognition result of 92% using linear discriminant analysis (LDA) was reported, but they only included nine people's face images and, hence, only 193 of the 213 images were used.

The experimental results clearly show that our approach outperforms the other methods. In addition, it should be noted that only two positions of the eyes are needed in our approach while 34 fiducial points are manually selected in other methods [3–5].

It should also be noted that in the JAFFE database, some expressions had been expressed inaccurately or labeled incorrectly. This also influences the recognition results. Fig. 3 shows a few examples with the labeled expression and our recognition results.

CONCLUSIONS

How to get a proper combination for face representation and classification is crucial for facial expression recognition. The combination of the local binary pattern operator and the linear programming technique is one solution to this question. The Local Binary Pattern operator, which has demonstrated excellent performance in texture classification and face recognition, is used in our method to describe a face efficiently for expression recognition. Then, 21 classifiers are produced based on linear programming technique and classification is implemented with a binary tree tournament scheme. Experimental results demonstrate that our method performs better than other methods on the JAFFE database. These preliminary results are promising and it is of interest to experiment with different facial expression databases.

REFERENCES

1. M. Pantic and L. J. M. Rothkrantz, "Automatic Analysis of Facial Expressions: The State of the Art," IEEE Trans. Pattern. Anal. Mach. Intell. **22** (12), 1424–1445 (2000).
2. B. Fasel and J. Luettin, "Automatic Facial Expression Analysis: A Survey," Pattern Recogn. **36** (1), 259–275 (2003).
3. Z. Zhang, "Feature-Based Facial Expression Recognition: Sensitivity Analysis and Experiment with a Multi-Layer Perceptron," Int. J. Pattern Recogn. Artif. Intell. **13** (6), 893–911 (1999).
4. G. D. Guo and C. R. Dyer, "Simultaneous Feature Selection and Classifier Training Via Linear Programming: A Case Study for Face Expression Recognition," in *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, 2003*, pp. 346–352.
5. M. Lyons, J. Budynek, and S. Akamastu, "Automatic Classification of Single Facial Images," IEEE Trans. Pattern. Anal. Mach. Intell. **21** (12), 1357–1362 (1999).
6. I. Buciu, C. Kotropoulos, and I. Pitas, "ICA and Gabor Representation for Facial Expression Recognition," in *Proceedings of Int. Conference on Image Processing, 2003*, pp. 855–858.
7. T. Oja, M. Pietikäinen, and T. Mäenpää, "Multiresolution Gray-Scale and Rotation Invariant Texture Classification with Local Binary Patterns," IEEE Trans. Pattern. Anal. Mach. Intell. **24** (7), pp. 971–987 (2002).
8. P. S. Bradley and O. L. Mangasarian, "Feature Selection via Concave Minimization and Support Vector Machines," in *Proceedings 5th Int. Conference Machine Learning, 1998*, pp. 82–90.
9. M. Lyons, S. Akamastu, M. Kamachi, and J. Gyoba, "Coding Facial Expressions with Gabor Wavelets," in *Proceedings of Third IEEE Conference Face and Gesture Recognition, 1998*, pp. 200–205.
10. D. Bolme, M. Teixeira, J. Beveridge, and B. Draper, "The CSU Face Identification Evaluation System User's Guide: Its Purpose, Feature and Structure," in *Proceedings of 3rd International Conference on Computer Vision Systems, 2003*, pp. 304–313.