Facial Expression Recognition Using Expression-Specific Local Binary Patterns and Layer Denoising Mechanism

[1] Wei-Lun Chao, [2] Jun-Zuo Liu, [3] Jian-Jiun Ding, [4] Po-Hung Wu
[1, 2, 3, 4] Graduate Institute of Communication Engineering
National Taiwan University
Taipei, Taiwan
[1] weilunchao@gmail.com, [2] jinzuo2011@gmail.com, [3] djj@ee.cc.ntu.edu.tw, [4] bz400@hotmail.com

Abstract—**In this paper, a novel framework for facial expression recognition is proposed, which improves the conventional feature extraction technique to further exploit distinctive characters for each label. To reduce the effect from unrelated features for facial expression recognition, a denoising mechanism is introduced. After denoising, to keep the connection between expression labels and whiten features as well as reduce the amount of computation, a manifold learning algorithm is applied, which finding a meaningful low-dimensional structure hidden in the whiten feature space. Finally, the features in the low-dimensional space are fed into the well know classifier such as the support vector machine and *k*-Nearest Neighbors. Simulations show that the proposed framework achieves the best recognition performance against existing methods in facial expression recognition.**

Keywords- facial expression recognition; local binary patterns; dimensionality reduction; manifold learning; machine learning

I. Introduction

Facial expression recognition and analysis (FERA) has attracted increasing attention in computer vision and pattern recognition fields because of its wide range of applications in human computer interface, public security, consumer electronics, clinic medicine, and so on. Over the past several decades, enormous efforts have been made and remarkable achievements are obtained in facial expression analysis[1][2][3]. However, facial expression recognition and analysis is still a challenging and complex task for computer to automatically determine expression due to the subtlety, complexity and variability of facial expressions.

Existing approaches of facial expression recognition can be categorized into video-based recognition model [4][5] and static image-based recognition model [6][7][8] in terms of original data source. The drawback of video-based model is the time-consumed computation due to the enormous data extracted from the sequence of images. With the speedy development of smart phones, a great number of real-time applications are designed under limited power consumption. As a result, using only one image to obtain enough information to determine the facial expression is the trend in the future. Based on these points, how to design the efficient facial expression recognition from static images is our concern in this paper.

The first step for facial expression recognition is feature extraction. It significantly affects the recognition performance. Although many approaches of feature extraction have been proposed to exploit the expression information [7][9], these approaches extract other information which is unrelated to facial expression as well. Therefore, the *expression-specific local binary pattern feature (es-LBP)*, which improves the conventional local binary patter feature (LBP) [15] such that the distinctive characters for each expression label can be further exploited, is proposed in this paper.

Another key issue discussed in our work is that the extracted features are very noisy, caused by the dimensions in the feature space that may be important for other recognition tasks such as identity recognition or gender recognition but not for facial expression recognition. To filter out the expression-specific feature dimensions such that in subsequent training phase the classifier can prevent from the effect of overfitting, a denoising mechanism is introduced. Another advantage of the denoising process is that only the expression-specific feature dimensions are picked up from the noisy feature space so that the resulting dimensionality of the denoised feature space is much lower than the vast one of the noisy feature space.

In some existing works [1][2], optical flow analysis has been used to model muscles activities or estimate the displacements of feature points. Facial geometry analysis has been widely exploited in facial representation [7], where shapes and locations of facial components are extracted to represent the face geometry. Another kind of method to represent faces is to model the appearance changes of faces. Holistic spatial analysis including principle component analysis (PCA) [6], linear discriminant analysis (LDA) [6], independent component analysis (ICA) [9], and Gabor wavelet analysis [9] have been applied to either the whole face or specific face regions to extract the facial appearance changes. Since the superior performance, Gabor wavelet representation has been widely used in face images analysis. However, its computation loading is large. Recently, Local Binary Patterns have been introduced as effective appearance features for facial image analysis.

This paper is organized as follows: In Section II, an overview of the proposed framework is presented and the concept of expression-specific local binary pattern feature and the denoising mechanism we proposed are introduced in Sections III and IV respectively. In Section V, the manifold learning algorithm for dimensionality reduction is demonstrated. The simulation results and the conclusion are presented in Section VI and Section VII.

II. Framework of the Proposed Algorithm

Given a training set $I = \{i^{(n)}\}_{n=1}^{N}$ with N facial images and its corresponding label set $Y = \{y^{(n)} \in L\}_{n=1}^{N}$, facial expression recognition can be modeled as a supervised learning task.

In this paper, a new facial expression recognition framework is proposed, which consist of four steps: Feature extraction, denoising mechanism, dimensionality reduction, and expression determination.

Improving the conventional Local Binary Patterns (LBP), the proposed expression-specific Local Binary Patterns (es-LBP) is adopted for feature extraction, which represent the input face image in a mathematic way and results in a feature vector $x \in R^D$ for each image i. Then, the proposed denoising mechanism is introduced to reduce the effect due to the noisy dimensions which are unrelated to facial expression recognition. The resulting features after these two steps are denoted as $x_{denoised} \in R^d$. For the dimensionality reduction step, Locality preserving projections algorithm is employed in advance to find the meaningful low dimension feature space R^p from the high dimension space R^d which result in a more impact representation $z \in R^p$. Finally, according to $z \in R^p$, an expression determination function is trained to classify the expression label $\hat{y}$. The flowchart of the proposed framework is shown in Fig. 1.

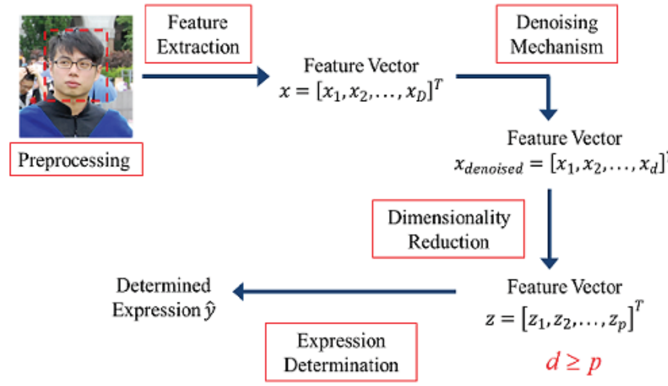


Figure 1. Flowchart of the proposed framework.

III. Feature Extraction

A. An Intorduction to the Original Local Binary Patterns

The original Local Binary Patterns operator, introduced by Ojala et al. [14], is a powerful method of texture description. The operator labels pixels of an image by thresholding a 3 × 3 neighborhood of each pixel with its pixel value and considering the result as a binary number. Then the histogram of labels (denoted by H_i) computed from an image $f_{l(x, y)}$ can be used as a texture descriptor. It is defined as:

$$H_i = \sum_{x,y} [[f_{l(x,y)} = i]], \; i = 0, \ldots, M-1 \tag{1}$$

where M is the number of different labels produced by the LBP operator. This histogram contains information about the distribution of the local micro-patterns, such as edges, spots and flat areas, over the whole image.

For efficient face representation, one should retain also spatial information. For this purpose, a face image should be divided into K regions, and the spatially enhanced histogram is defined as:

$$H_{i,j} = \sum_{x,y} [[f_{l(x,y)} = i]] \, [[(x, y) \in R_j]] \tag{2}$$

where $i = 0, 1, \ldots, M-1$ and $j = 0, 1, \ldots, K-1$.

B. The Expression-specific Local Binary Patterns

The proposed expression-specific Local Binary Patterns (es-LBP) consist of three parts: The large-scale face image, the small-scale face image and the interesting feature regions as demonstrated in Fig. 2. The concept of the expression-specific Local Binary Patterns is derived from the phenomenon we observed where human beings can determine a facial expression not only from the situation that the human face is very close but also from the situation that the face is very far away. Moreover, even though there are some obstacles to the face, the human beings still can detect exactly the correct expression. That means human beings can recognize the facial expression according to the certain feature regions of the face but not to the whole face. In this paper, there are eight interesting feature regions: forehead region, right eyebrow region, left eyebrow region, right eye region, left eye region, right cheek region, left cheek region, and mouth region.

The first part of es-LBP, the large-scale face image, is analog to the situation that the human face is very close and the second part of es-LBP, the small-scale face image, corresponds to the situation that the human face is far away. For the third part of es-LBP (the interesting feature regions), another factor, which is jointly considered with the situation that there are some obstacles to the face, is that the original LBP did not take the symmetry of the expression display from the interesting feature region into consideration.

Take the eyebrow region for example. If the original LBP operator is applied directly, the histogram representation of the eyebrow from the happy face would be the same as the one from the sad face shown in Fig. 3.

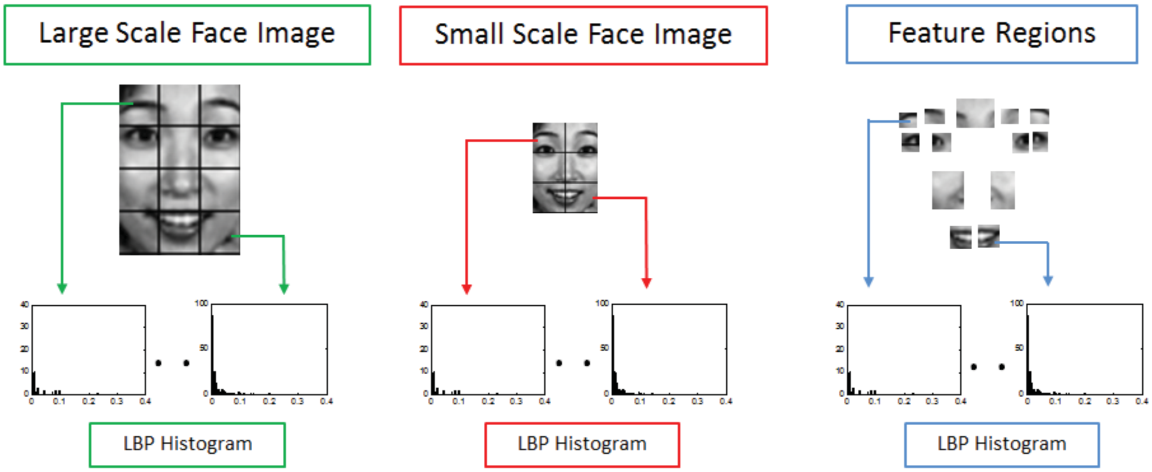


Figure 2. Demonstration of the proposed es-LBP feature including three parts, namely large scale face image, small scale face image, and feature regions from left to right respectively.

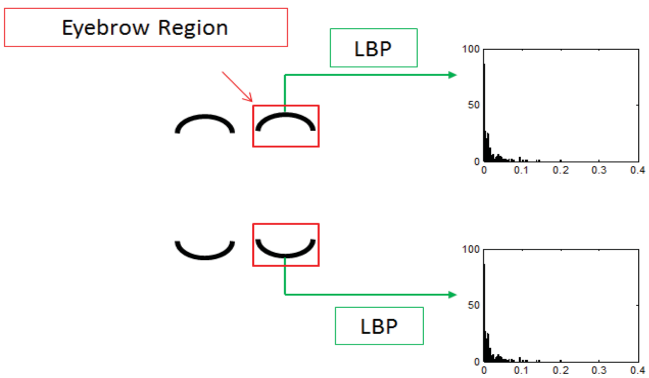


Figure 3. The two LBP histogram representations from the eyebrow of happy face and sad face would be the same since the original LBP did not take the symmetry of patters into consideration.

The original LBP operator statistically computes the distribution of the local micro-patterns, such as edges, spots, and flat areas and categorizes these patterns into the histogram bins separately. From the statistic view, a histogram representation could be taken as the compound of many patterns with different frequency and the summation of these patterns frequency consistent with the histogram representation. For this reason, the patterns related to expression recognition may be jointly mixed with the patterns unrelated to expression recognition to form the histogram representation. Consequently, the original LBP operator cannot access the patterns related to expression recognition exclusively. The third part of es-LBP (the interesting feature regions) resolves this problem. The interesting feature regions are divided into two parts respectively and the LBP operator is applied to the divided parts separately. By this procedure, the statistical property of the expression-specific patterns can be accessed.

IV. DENOISING MECHANISM

After feature extraction, the feature vectors are very noisy, caused by the dimensions in the feature space that may be important for other recognition tasks such as identity recognition or gender recognition but not for facial expression recognition. To filter out the expression-specific feature dimensions such that in subsequent training phase the classifier can prevent from the effect of overfitting, a denoising mechanism is introduced. The three parts of es-LBP features are denoised severally. The reason is that each part of the es-LBP feature is viewed as a special case on which the expression can be determined based. Based on this reason, the feature dimensions (unrelated to expression recognition) existing in the three feature spaces, expanded by the three parts of es-LBP feature, are different. Therefore, the denoising mechanism should be applied to each feature space respectively. The technique used in denoising mechanism process is the well-known Principle Component Analysis (PCA) [12]. PCA is a kind of feature transformation that preserves maximum variance while reducing the unimportant dimensions as more as possible. Another advantage of the denoising process is that only the expression-specific feature dimensions are picked up from the noisy feature space so that the resulting dimensionality of the denoised feature space is much lower than the vast one of the noisy feature space.

V. DIMENSIONALITY REDUCTION

In order to increase the efficiency, not only the denoising mechanism is proposed to reduce the amount of data but also dimensionality reduction is applied. There are many nonlinear dimensionality reduction algorithms have been investigated in the past decade, for example, Locally Linear Embedding (LLE) was proposed by S. T. Roweis et al. [13] and Locality Preserving Projection (LPP) was proposed by X. He et al. [11]. In this paper, another dimension reduction algorithm, Marginal Fisher Analysis (MFA) [16] is applied to support the generality of the proposed es-LBP and the denoising mechanism. The idea of Marginal Fisher Analysis (MFA) is based on Linear Discriminant Analysis (LDA) and seeks to preserve neighbors with the same label while pull away neighbors with different labels. The intrinsic graph in MFA characterizes the intra-class compactness and connects each sample with its neighbor samples of the same class, while penalty graph connects the marginal samples and characterizes the interclass separability.

VI. Simulation Results

The proposed facial expression recognition is evaluated on the widely-used JAFFE database which consists of 213 images of Japanese female facial expression. There are 3 or 4 images for each of the seven basic expressions (angry, despise, fear, happy, neutral, sad, surprise) and all 213 images of the JAFFE database are used for 7-class expression recognition. The preprocessing for the database in our work includes only face detection and eye detection, no additional step is applied such as face alignment or illumination normalization. The frequently used face detection method, Viola and Jones's method, is used and the eye detection is accomplished by the source code in matlab central which is designed based on Viola and Jones's face detection. The detected face images are cropped to size of 85x60 for large-scale images, and linearly scaled to size 40x30 for small-scale images. The right eyebrow region, left eyebrow region, right eye region, left eye region, and mouth region are further divided into two sub regions respectively to obtain totally 13 sub-interesting feature regions.

Two popular measures, person dependent average accuracy and person independent average accuracy are used for evaluation. For person independent experiment, the leave-one-person-out (LOPO) testing strategy is adopted, where the estimation algorithm is repeatedly trained on images from 9 people and tested on images of the remaining person for 10 times. On the other hand, the 10-fold cross-validation is adopted for person independent experiment, where the database is randomly partitioned into 10 parts, select nine parts for training and test the recognition rate on the remaining part. Repeat above process for 10 times such that every part is taken as the testing part once. Finally, the average of recognition rate is calculated to identify the performance of algorithm.

In our work, two well-known classify algorithm (Support Vector Machine [10] and *K*-Nearest Neighborhood) are chosen as final classifiers. To show the efficacy of the proposed es-LBP feature and denoising mechanism, two parts experiments are executed respectively. The first part is to explore whether the proposed es-LBP is better than the original LBP feature, and the second part is to verify the necessity of denoising mechanism. In the first part, experiments adopted the conventional recognition framework which includes three steps (feature extraction, dimensionality reduction and expression determination). The results of the first part experiment are listed in Tables I-IV.

From Table I, the person dependent average accuracy with the proposed es-LBP feature is better than the one with original LBP feature in the LPP vector space, no matter which classifier is applied, as well as in the MFA vector space shown in Table II. Based on the results, given the same setting of other steps in the proposed framework, using the proposed es-LBP feature always achieve better recognition performance.

For person independent experiments, as listed in Tables III-IV, the experimental results show that using the proposed es-LBP with 13 features indeed improve the recognition performance, which is about 3%~5% better than the one of original LBP feature.

Table I. Person Dependent Average Accuracies Based on the LPP Dimensionality Reduction Algorithm

	KNN	SVM
LBP	0.8619	0.8810
es-LBP with 8 features	0.8762	0.8905
es-LBP with 13 features	0.8905	0.8952

Table II. Person Dependent Average Accuracy Based on the MFA Dimensionality Reduction Algorithm

	KNN	SVM
LBP	0.8762	0.8810
es-LBP with 8 features	0.9048	0.9143
es-LBP with 13 features	0.9143	0.9190

Table III. Person Independent Average Accuracies Based on the LPP Dimensionality Reduction Algorithm.

	KNN	SVM
LBP	0.6048	0.6333
es-LBP with 8 features	0.6190	0.6429
es-LBP with 13 features	0.6381	0.6910

Table IV. Person Independent Average Accuracies Based on the MFA Dimensionality Reduction Algorithm

	KNN	SVM
LBP	0.6286	0.6429
es-LBP with 8 features	0.6381	0.6571
es-LBP with 13 features	0.6524	0.6905

Table V. Person Dependent Average Accuracies Based on the LPP between the Algorithms with and without Denoising Mechanism

	Non-Denoising	Denoising
LBP	0.8619	0.8762
es-LBP with 8 features	0.8762	0.9000
es-LBP with 13 features	0.8905	0.9143

Table VI. Comparison between the Proposed Approach and the Existing Methods

Method	Feature extraction	Dimension reduction	Classifier	LOPO	10-fold cross validation
[6]	LBP	PCA	SVM	53.8	N/A
[6]	LBP	KPCA	SVM	53.6	N/A
[6]	LBP	LDA	SVM	55.7	N/A
[8]	Gabor filter	N/A	Linear Programing	N/A	91.0
[9]	Gabor filter	N/A	SVM	N/A	90.34
[9]	ICA	N/A	SVM	N/A	79.91
[7]	AAM	N/A	SVM	N/A	89.5
Proposed	**es-LBP**	**MFA**	**SVM**	**65.24**	**91.9**

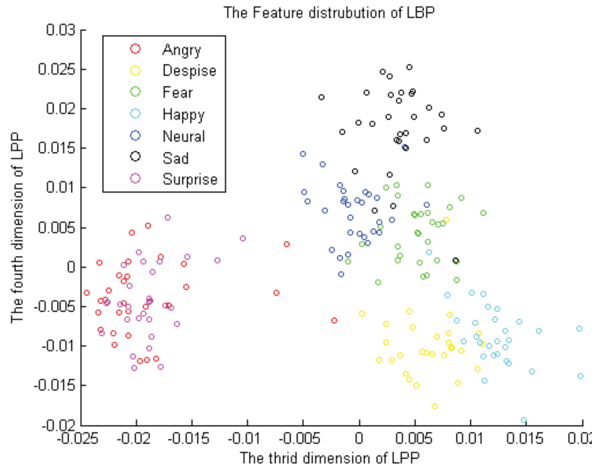
In the second part experiments, the denoising mechanism is performed before the dimensionality reduction step in the first part experiments to show its effect on recognition performance. As shown in Table V, the denoising mechanism increases about 1~3% recognition rate, as a result, performing the proposed denoising mechanism really can improves the conventional recognition framework. To further ensure that the proposed es-LBP exploit distinctive characters for each label, the first two features in LPP vector space of the whole database are visualized in Fig. 4. Obviously, LBP features with different label are mixed together and cannot be separate smoothly, while the between-class scatter of the es-LBP features are more larger which is desiring thing in recognition task. Finally, the comparison between the proposed approach and the existing methods is shown in Table VI.

VII. Conclusion

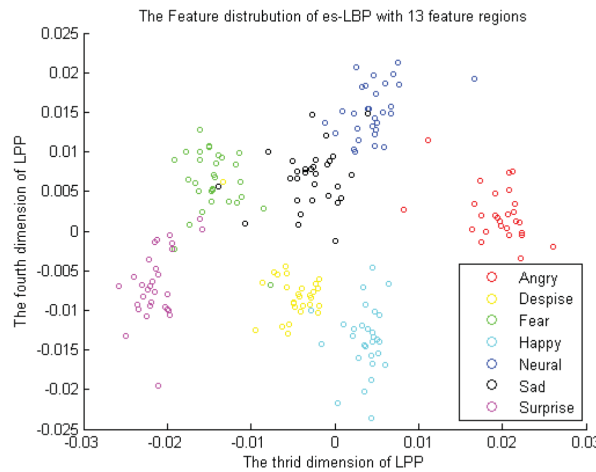
A new facial expression recognition framework which includes four steps (feature extraction, denoising mechanism, dimensionality reduction, and expression determination) is proposed. The proposed es-LBP features, which modify the conventional LBP features, consider the symmetry of histogram patterns and further exploits distinctive characters for each label are shown to be more suitable for facial expression recognition than original LBP. In addition, the necessity of the proposed denoising mechanism which reduces the effect from unrelated features for facial expression recognition is verified to be important in facial recognition framework. From the simulation results, the proposed approach achieves the better performance against the state-of-art methods

Acknowledgment

This work was supported by the National Science Council, R.O.C., under the contract 100-2221-E-002-206-MY3.



(a)



(b)

Figure 4. Feature distributions of (a) LBP and (b) es-LBP in the third and fourth dimension LPP vector space.

References

[1] M. Pantic and L. Rothkrantz, "Automatic analysis of facial expressions: The state of art," IEEE Trans. Pattern Analysis and Machine Intelligence, vol. 22, issue 6, pp. 1424-1445, Dec. 2000.

[2] B. Fasel and J. Luettin, "Automatic facial expression analysis: A survey," Pattern Recognition, vol. 36, issue 1, pp. 259-275, Jan. 2003.

[3] M. Valstar, B. Jiang, M. Mehu, M. Pantic, and K. Scherer, "The first facial expression recognition and analysis challenge," Automatic Face & Gesture Recognition and Workshop, pp. 921-926, Mar. 2011.

[4] I. Kotsia and I. Pitas, "Facial expression recognition in image sequences using geometric deformation features and support vector machines," ICIP, vol. 16, issue 1, Jan. 2007.

[5] G. Zhao and M. Pietikainen, "Dynamic texture recognition using local binary patterns with an application to facial expressions," IEEE Trans. Pattern Analysis and Machine Intelligence, vol. 29, issue. 6, pp. 915-928, Jun. 2007.

[6] Y. Zilu and F. Xieyan, "Combining LBP and Adaboost for facial expression recognition," International Conference on Signal Processing, pp. 1461-1464, Oct. 2008.

[7] G. Lei, X. H. Li, J. L. Zhou, and X. G. Gong, "Geometric feature based facial expression recognition using multiclass support vector machines," IEEE International Conference on Granular Computing, pp. 318-321, Aug. 2009.

[8] G. Guo and C. Dyer, "Simultaneous feature and classifier training via linear programming: A case study for face expression recognition," Computer Vision and Pattern Recognition Workshop, vol. 1, pp. 346-352, Jun. 2003.

[9] I. Buciu, C. Kotropoulos, and I. Pitas, "ICA and Gabor representation for facial expression recognition," International Conference on Image Processing, vol. 2, pp. 855-858, Sept. 2003.

[10] C. C. Chang and C. J. Lin., "LIBSVM: A library for support vector machines," Software available at http://www.csie.ntu.edu.tw~cjlin/libsvm, 2001.

[11] Xiaofei He and Partha Niyogi, Advances in Neural Information Processing System 16, MIT Press, 2004.

[12] M. E. Tipping and C. M. Bishop, "Probabilistic Principle Component Analysis," Journal of the Royal Statistical Society, vol. 6, no. 3, pp. 611-622, 1999.

[13] S. T. Roweis and L. K. Saul, "Nonlinear dimensionality reduction by locally linear embedding," Science, vol. 290, no. 5500, pp. 2323-2326, Dec. 2000.

[14] P. Viola and M. Jones, "Rapid object detection using a boosted cascade of simple features," IEEE Conference on Computer Vision and Pattern Recognition, vol. 1, pp. 511-518, 2007.

[15] T. Ahonen, A. Hadid, and M. Pietikainen, "Face description with local binary patterns: Application to face recognition," IEEE Trans. Pattern Analysis and Machine Intelligence, vol. 28, issue. 12, pp. 2037-2041, Dec. 2006.

[16] D. Xu, et al., "Marginal Fisher analysis and its variants for human gait recognition and content-based image retrieval," ICIP, vol. 16, issue. 11, pp. 2811-2821, Nov. 2007.